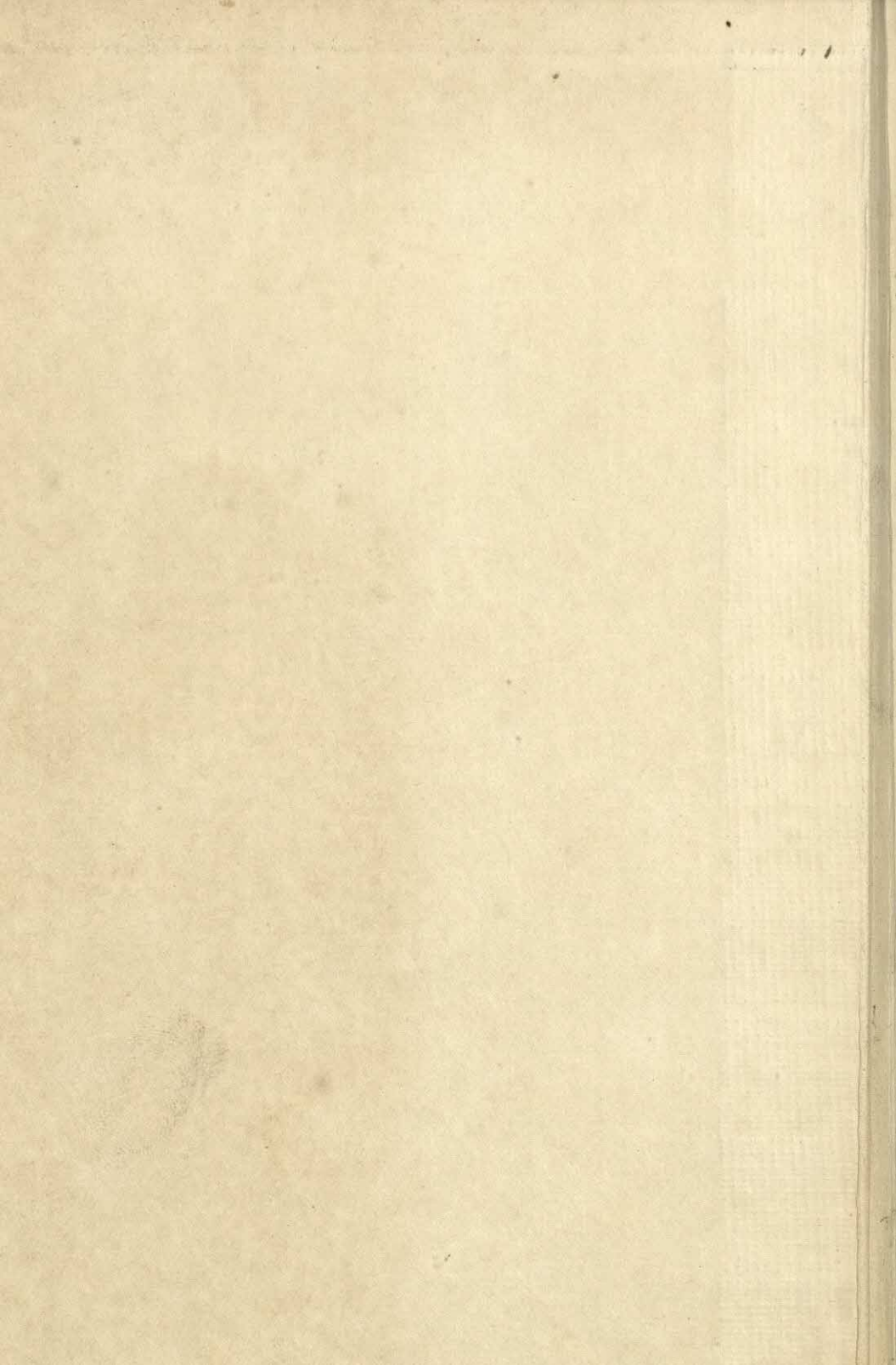
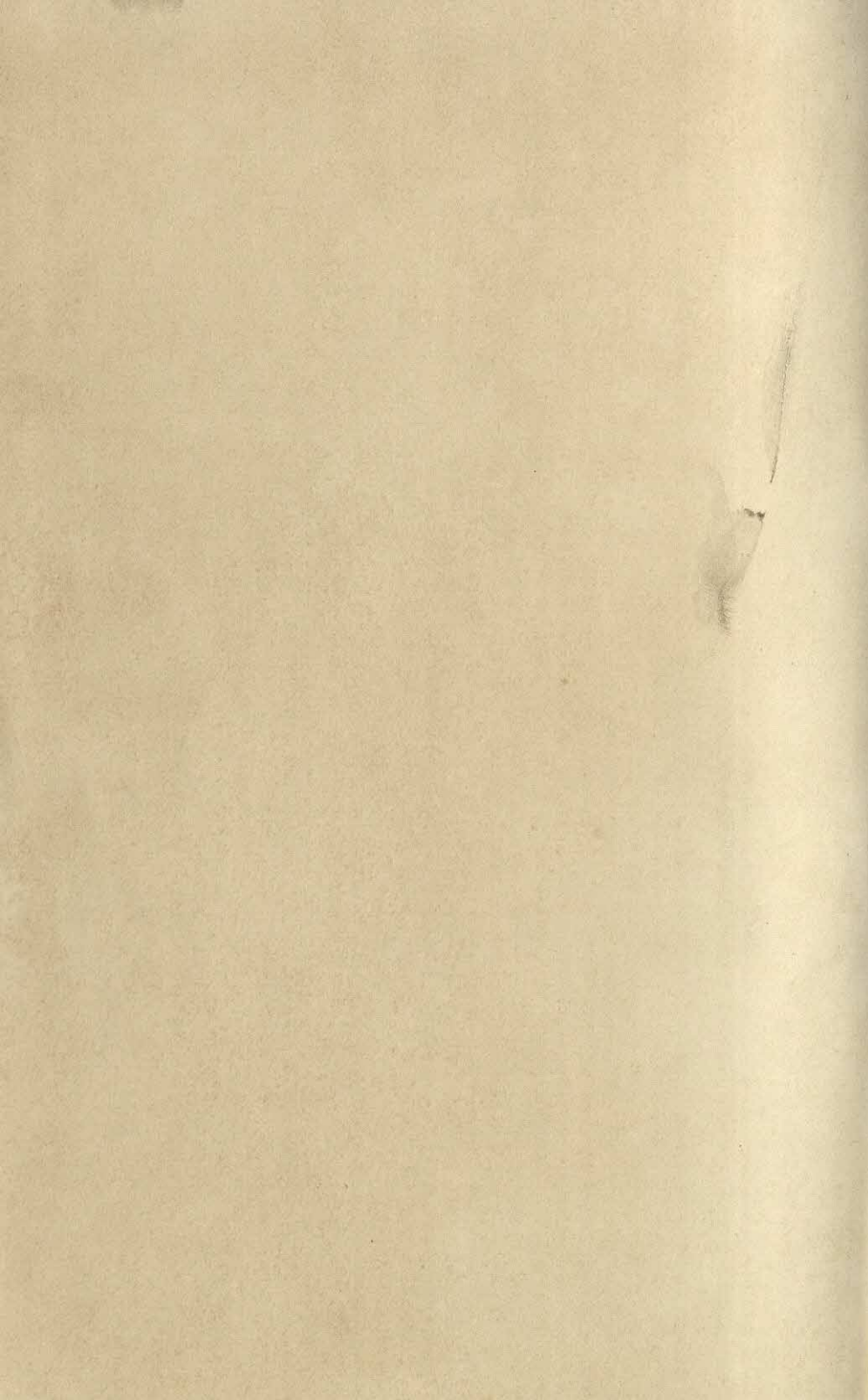
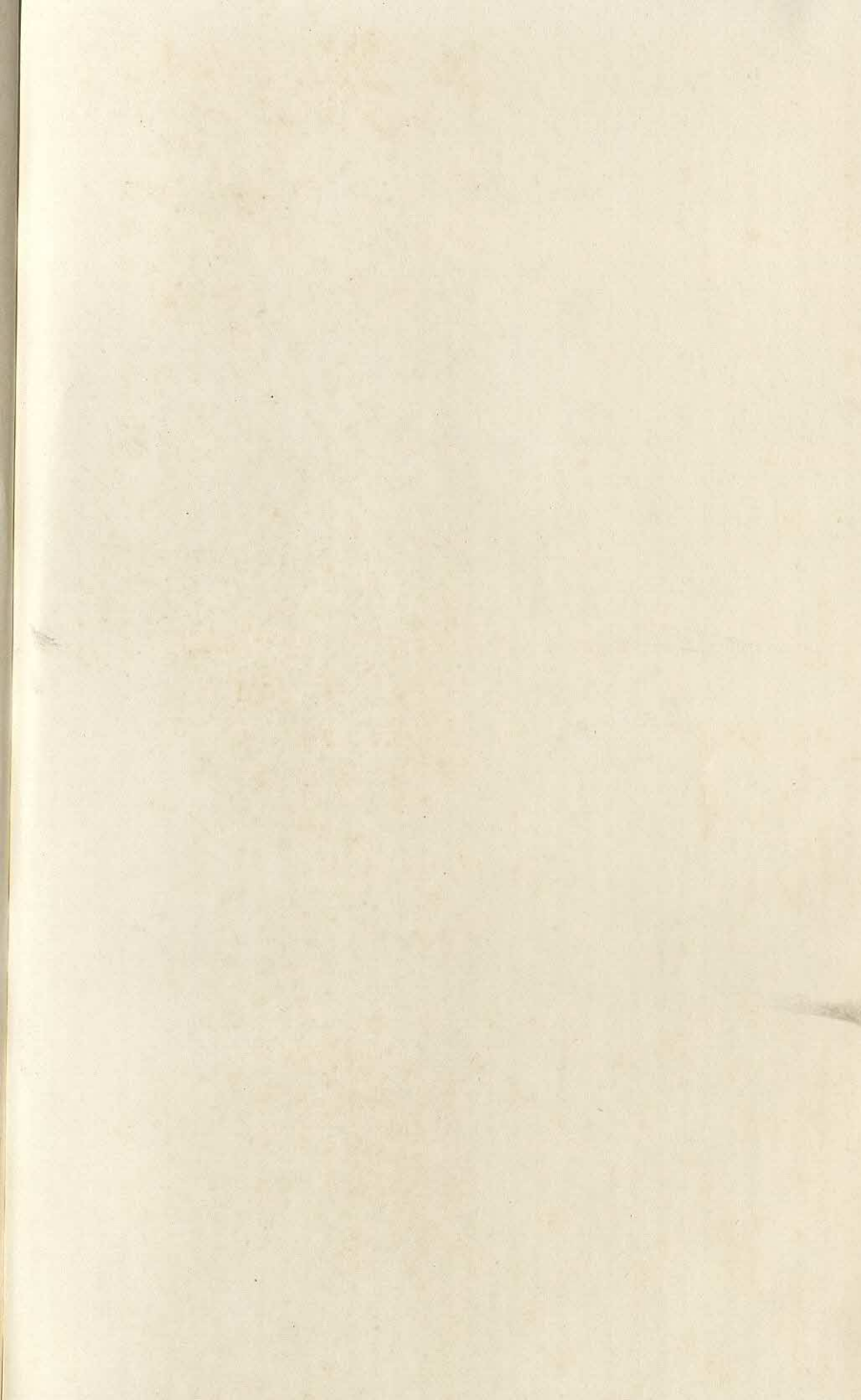
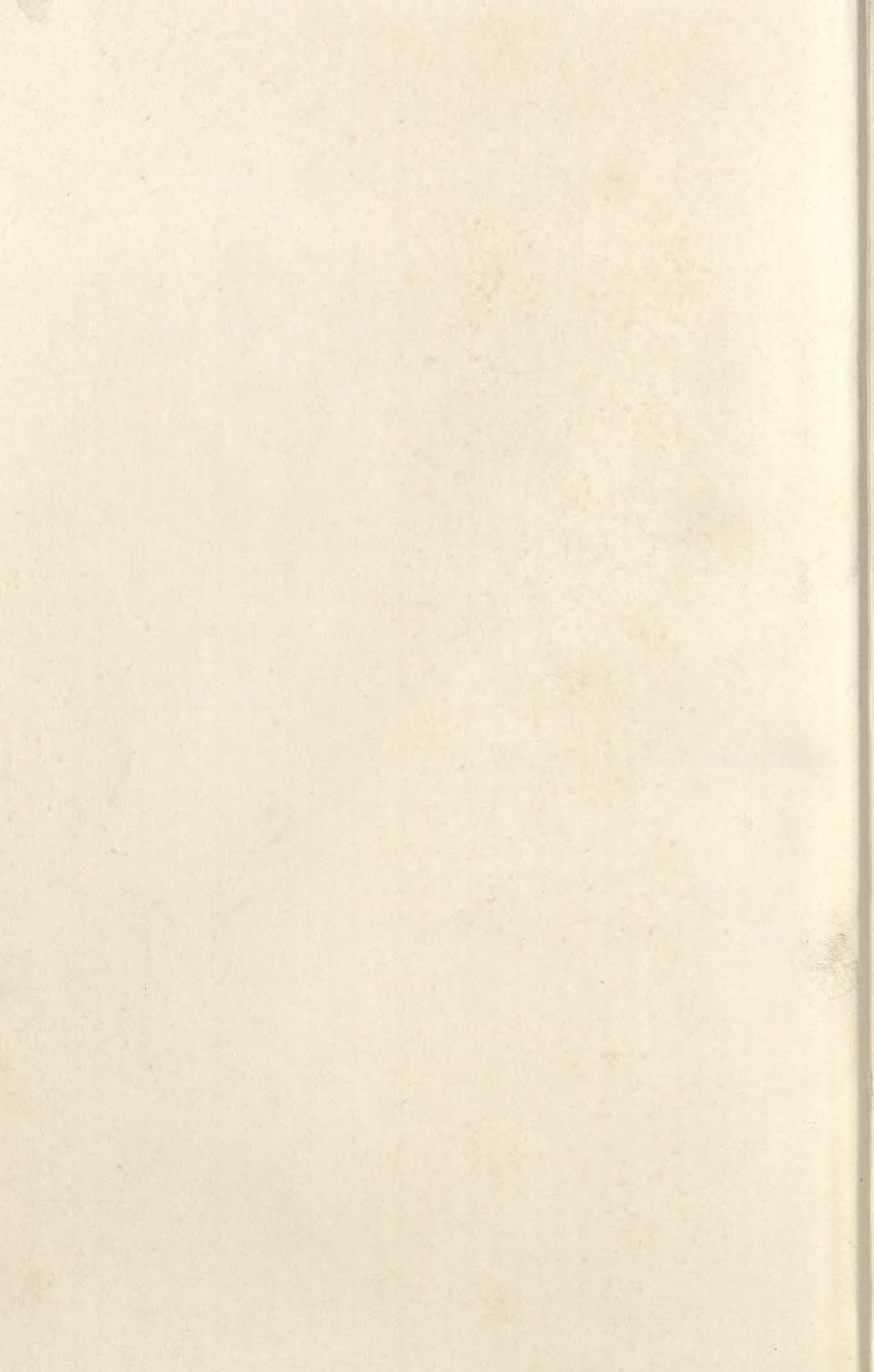

McGraw-Hill
Series in
Psychology



1766.







Statistical Principles in Experimental Design

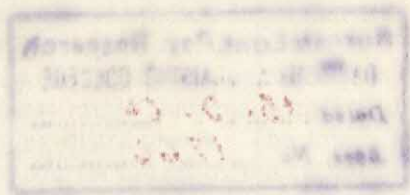
McGRAW-HILL SERIES IN PSYCHOLOGY

HARRY F. HARLOW, *Consulting Editor*

BEACH, HEBB, MORGAN, AND NISSEN · The Neuropsychology of Lashley
VON BÉKÉSY · Experiments in Hearing
BERLYNE · Conflict, Arousal, and Curiosity
BLUM · Psychoanalytic Theories of Personality
BROWN · The Motivation of Behavior
BROWN · The Psychodynamics of Abnormal Behavior
BROWN AND GHISELLI · Scientific Method in Psychology
CATTELL · Personality
COFER · Verbal Learning and Verbal Behavior
CRAFTS, SCHNEIRLA, ROBINSON, AND GILBERT · Recent Experiments in Psychology
DEESE · The Psychology of Learning
DOLLARD AND MILLER · Personality and Psychotherapy
DORCUS AND JONES · Handbook of Employee Selection
FERGUSON · Personality Measurement
FERGUSON · Statistical Analysis in Psychology and Education
GHISELLI AND BROWN · Personnel and Industrial Psychology
GILMER · Industrial Psychology
GRAY · Psychology Applied to Human Affairs
GUILFORD · Fundamental Statistics in Psychology and Education
GUILFORD · Psychometric Methods
GUILFORD · Personality
HAIRE · Psychology in Management
HIRSH · The Measurement of Hearing
HURLOCK · Adolescent Development
HURLOCK · Child Development
HURLOCK · Development Psychology
KARN AND GILMER · Readings in Industrial and Business Psychology
KRECH AND CRUTCHFIELD · Theory and Problems of Social Psychology
LAZARUS · Adjustment and Personality
LEWIN · A Dynamic Theory of Personality
LEWIN · Principles of Topological Psychology
LEWIS · Quantitative Methods in Psychology
MAIER AND SCHNEIRLA · Principles of Animal Psychology
MILLER · Language and Communication
MISIAK AND STAUDT · Catholics in Psychology: A Historical Survey
MOORE · Psychology for Business and Industry
MORGAN AND STELLAR · Physiological Psychology
PAGE · Abnormal Psychology
REYMERT · Feelings and Emotions
SEASHORE · Psychology of Music
SHAFFER AND LAZARUS · Fundamental Concepts in Clinical Psychology
SIEGEL · Nonparametric Statistics: for the Behavioral Sciences
STAGNER · Psychology of Personality
TOWNSEND · Introduction to Experimental Method
VINACKE · The Psychology of Thinking
WALLEN · Clinical Psychology: The Study of Persons
WATERS, RETHLINGSHAFFER, AND CALDWELL · Principles of Comparative Psychology
WINER · Statistical Principles in Experimental Design
ZUBEK AND SOLBERG · Human Development

John F. Dashiell was Consulting Editor of this series from its inception in 1931 until January 1, 1950. Clifford T. Morgan was Consulting Editor of this series from January 1, 1950 until January 1, 1959.

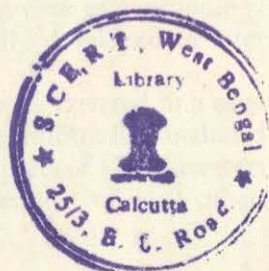
STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN



1c
B. J. WINER

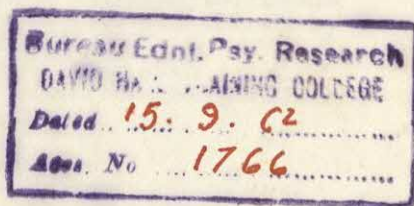
Professor of Psychology and Statistics
Purdue University

McGRAW-HILL BOOK COMPANY, INC. 1962
New York San Francisco Toronto London



U.E.R.T. West Bengal
Date 15.9.62
Acc. No. 1766

150
WIN



STATISTICAL PRINCIPLES IN EXPERIMENTAL DESIGN

Copyright © 1962 by the McGraw-Hill Book Company, Inc. Printed in the United States of America. All rights reserved. This book, or parts thereof, may not be reproduced in any form without permission of the publishers. *Library of Congress Catalog Card Number 61-13174*

70980



Preface

Written primarily for students and research workers in the area of the behavioral sciences, this book is meant to provide a text and comprehensive reference source on statistical principles underlying experimental design. Particular emphasis is given to those designs that are likely to prove useful in research in the behavioral sciences.

The book primarily emphasizes the logical basis of principles underlying designs for experiments rather than mathematical derivations associated with relevant sampling distributions. The topics selected for inclusion are those covered in courses taught by the author during the past several years.

Students in these courses have widely varying backgrounds in mathematics and come primarily from the fields of psychology, education, economics, sociology, and industrial engineering. It has been the intention of the author to keep the book at a readability level appropriate for students having a mathematical background equivalent to freshman college algebra. From experience with those sections of the book which have been used as text material in dittoed form, there is evidence to indicate that, in large measure, the desired readability level has been attained. Admittedly, however, there are some sections in the book where this readability goal has not been achieved.

The first course in design, as taught by the author, has as a prerequisite a basic course in statistical inference. The contents of Chaps. 1 and 2 review the highlights of what is included in the prerequisite material. These chapters are not meant to provide the reader with a first exposure to these topics. They are intended to provide a review of terminology and notation for the concepts which are more fully developed in later chapters.

By no means is all the material included in the book covered in a one-semester course. In a course of this length, the author has included Chaps. 3, 4, parts of 5, 6, parts of 7, parts of 10, and parts of 11. Chapters 8 through 11 were written to be somewhat independent of each other.

Hence one may read, with understanding, in these chapters without undue reference to material in the others.

In general, the discussion of principles, interpretations of illustrative examples, and computational procedures are included in successive sections within the same chapter. However, to facilitate the use of the book as a reference source, this procedure is not followed in Chaps. 5 and 6. Basic principles associated with a large class of designs for factorial experiments are discussed in Chap. 5. Detailed illustrative examples of these designs are presented in Chap. 6. For teaching purposes, the author includes relevant material from Chap. 6 with the corresponding material in Chap. 5.

Selected topics from Chaps. 7 through 11 have formed the basis for a second course in experimental design.

Relatively complete tables for sampling distributions of statistics used in the analysis of experimental designs are included in the Appendix. Ample references to source materials having mathematical proofs for the principles stated in the text are provided.

The author is indebted to E. S. Pearson and the trustees of *Biometrika* for permission to reproduce parts of Tables B.1, B.3, B.7, and B.9 from *Biometrika Tables for Statisticians*, vol. I, 2d ed. The author is indebted to H. L. Harter, D. S. Clem, and E. H. Guthrie for permission to reproduce Table B.4, which was taken from WADC Technical Report 58-484, vol. II, 1959. The author is indebted to C. W. Dunnett and the editor of the *Journal of the American Statistical Association* for permission to reprint Table B.6. The author is also indebted to C. Eisenhart, M. W. Hastay, and W. A. Wallis for permission to reprint Table B.8, which appears in *Techniques of Statistical Analysis*, 1947. The author is also indebted to L. S. Feldt and M. W. Mahmoud as well as the editor of *Psychometrika* for permission to reprint Table B.11.

Special thanks are due to Mrs. G. P. Lehman and Mrs. R. L. Smith for excellent secretarial assistance in preparing the manuscript.

The author is particularly grateful to Dr. D. A. Wood for many reasons, and to Dr. A. Lubin, whose critical reading of the manuscript did much to help the author prepare the present version of this book.

B. J. Winer

Contents

<i>Preface</i>	v
<i>Introduction</i>	1
CHAPTER 1. <i>Basic Concepts in Statistical Inference</i>	4
1.1 Basic terminology in sampling	4
1.2 Basic terminology in statistical estimation	6
1.3 Basic terminology in testing statistical hypotheses	9
CHAPTER 2. <i>Testing Hypotheses about Means and Variances</i>	14
2.1 Testing hypotheses on means— σ assumed known	14
2.2 Tests of hypotheses on means— σ estimated from sample data	20
2.3 Testing hypotheses about the difference between two means— assuming homogeneity of variance	24
2.4 Computational formulas for the t statistic	31
2.5 Test for homogeneity of variance	33
2.6 Testing hypotheses about the difference between two means— assuming that population variances are not equal	36
2.7 Testing hypotheses about the difference between two means— correlated observations	39
2.8 Combining several independent tests on the same hypothesis	43
CHAPTER 3. <i>Design and Analysis of Single-factor Experiments</i>	46
3.1 Introduction	46
3.2 Definitions and numerical example	48
3.3 Structural model for single-factor experiment—model I	56
3.4 Structural model for single-factor experiment—model II (variance component model)	62
3.5 Methods for deriving estimates and their expected values	63
3.6 Comparisons among treatment means	65
3.7 Use of orthogonal components in tests for trend	70
3.8 Use of the studentized range statistic	77
3.9 Alternative procedures for making a posteriori tests	85
3.10 Comparing all means with a control	89
3.11 Tests for homogeneity of variance	92
3.12 Unequal sample sizes	96
3.13 Determination of sample size	104

CHAPTER 4. <i>Single-factor Experiments Having Repeated Measures on the Same Elements</i>	105
4.1 Purpose	105
4.2 Notation and computational procedures	106
4.3 Numerical example	111
4.4 Statistical basis for the analysis	116
4.5 Use of analysis of variance to estimate reliability of measurements	124
4.6 Tests for trend	132
4.7 Analysis of variance for ranked data	136
4.8 Dichotomous data	138
CHAPTER 5. <i>Design and Analysis of Factorial Experiments</i>	140
5.1 General purpose	140
5.2 Terminology and notation	141
5.3 Main effects	146
5.4 Interaction effects	148
5.5 Experimental error and its estimation	150
5.6 Estimation of mean squares due to main effects and interaction effects	151
5.7 Principles for constructing F ratios	160
5.8 Higher-order factorial experiments	162
5.9 Estimation and tests of significance for three-factor experiments	170
5.10 Simple effects and their tests	174
5.11 Geometric interpretation of higher-order interactions	178
5.12 Nested factors (hierarchal designs)	184
5.13 Split-plot designs	191
5.14 Rules for deriving the expected values of mean squares	195
5.15 Quasi F ratios	199
5.16 Preliminary tests on the model and pooling procedures	202
5.17 Individual comparisons	207
5.18 Partition of main effects and interaction into trend components	211
5.19 Replicated experiments	213
5.20 The case $n = 1$ and a test for nonadditivity	216
5.21 The choice of a scale of measurement and transformations	218
5.22 Unequal cell frequencies	222
5.23 Unequal cell frequencies—least-squares solution	224
CHAPTER 6. <i>Factorial Experiments—Computational Procedures and Numerical Examples</i>	228
6.1 General purpose	228
6.2 $p \times q$ factorial experiment having n observations per cell	228
6.3 $p \times q$ factorial experiment—unequal cell frequencies	241
6.4 Effect of scale of measurement on interaction	244
6.5 $p \times q \times r$ factorial experiment having n observations per cell	248
6.6 Computational procedures for nested factors	258
6.7 Factorial experiment with a single control group	263
6.8 Test for nonadditivity	267
6.9 Computation of trend components	273
6.10 General computational formulas for main effects and interactions	278
6.11 Missing data	281

6.12	Special computational procedures when all factors have two levels	283
6.13	Illustrative applications	287
6.14	Unequal cell frequencies—least-squares solution	291
CHAPTER 7. <i>Multifactor Experiments Having Repeated Measures on the Same Elements</i>		
		298
7.1	General purpose	298
7.2	Two-factor experiment with repeated measures on one factor	302
7.3	Three-factor experiment with repeated measures (case I)	319
7.4	Three-factor experiment with repeated measures (case II)	337
7.5	Other multifactor repeated-measure plans	349
7.6	Tests on trends	353
7.7	Testing equality and symmetry of covariance matrices	369
7.8	Unequal group size	374
CHAPTER 8. <i>Factorial Experiments in Which Some of the Interactions Are Confounded</i>		
		379
8.1	General purpose	379
8.2	Modular arithmetic	381
8.3	Revised notation for factorial experiments	383
8.4	Method for obtaining the components of interactions	384
8.5	Designs for $2 \times 2 \times 2$ factorial experiments in blocks of size 4	394
8.6	Simplified computational procedures for 2^k factorial experiments	399
8.7	Numerical example of $2 \times 2 \times 2$ factorial experiment in blocks of size 4	404
8.8	Numerical example of $2 \times 2 \times 2$ factorial experiment in blocks of size 4 (repeated measures)	409
8.9	Designs for 3×3 factorial experiments	412
8.10	Numerical example of 3×3 factorial experiment in blocks of size 3	418
8.11	Designs for $3 \times 3 \times 3$ factorial experiments	423
8.12	Balanced $3 \times 2 \times 2$ factorial experiment in blocks of size 6	433
8.13	Numerical example of $3 \times 2 \times 2$ factorial experiment in blocks of size 6	439
8.14	$3 \times 3 \times 3 \times 2$ factorial experiment in blocks of size 6	442
8.15	Fractional replication	447
CHAPTER 9. <i>Balanced Lattice Designs and Other Balanced Incomplete-block Designs</i>		
		456
9.1	General purpose	456
9.2	Balanced simple lattice	457
9.3	Numerical example of balanced simple lattice	463
9.4	Balanced lattice-square designs	468
9.5	Balanced incomplete-block designs	477
9.6	Numerical example of balanced incomplete-block design	486
9.7	Youden squares	492
9.8	Numerical example of Youden square	496
9.9	Partially balanced designs	501
9.10	Numerical example of partially balanced design	506
9.11	Linked paired-comparison designs	511

CHAPTER 10. <i>Latin Squares and Related Designs</i>	514
10.1 Definition of Latin square	514
10.2 Enumeration of Latin squares	516
10.3 Structural relation between Latin squares and three-factor factorial experiments	519
10.4 Uses of Latin squares	521
10.5 Analysis of Latin-square designs—no repeated measures	524
10.6 Analysis of Greco-Latin squares	536
10.7 Analysis of Latin squares—repeated measures	538
CHAPTER 11. <i>Analysis of Covariance</i>	578
11.1 General purpose	578
11.2 Single-factor experiments	581
11.3 Numerical example of single-factor experiment	588
11.4 Factorial experiment	595
11.5 Computational procedures for factorial experiment	599
11.6 Factorial experiment—repeated measures	606
11.7 Multiple covariates	618
APPENDIX A. <i>Topics Closely Related to the Analysis of Variance</i>	622
A.1 Kruskal-Wallis H test	622
A.2 Contingency table with repeated measures	623
A.3 Comparing treatment effects with a control	627
A.4 General partition of degrees of freedom in a contingency table	629
A.5 Hotelling's T^2 test for the equality of k means	632
A.6 Least-squares estimators—general principles	635
APPENDIX B. <i>Tables</i>	639
B.1 Unit normal distribution	640
B.2 Student's t distribution	641
B.3 F distribution	642
B.4 Distribution of the studentized range statistic	648
B.5 Arcsin transformation	650
B.6 Distribution of t statistic in comparing treatment means with a control	652
B.7 Distribution of F_{\max} statistic	653
B.8 Critical values for Cochran's test for homogeneity of variance	654
B.9 Chi-square distribution	655
B.10 Coefficients of orthogonal polynomials	656
B.11 Curves of constant power for the test on main effects	657
B.12 Random permutations of 16 numbers	659
<i>Content References</i>	661
<i>References to Experiments</i>	665
<i>Index</i>	667

Introduction

The design of an experiment may be compared to an architect's plans for a structure, whether it be a giant skyscraper or a modest home. The basic requirements for the structure are given to the architect by the prospective owner. It is the architect's task to fill these basic requirements; yet the architect has ample room for exercising his ingenuity. Several different plans may be drawn up to meet all the basic requirements. Some plans may be more costly than others; given two plans having the same cost, one may offer potential advantages that the second does not.

In the design of an experiment, the designer has the role of the architect, the experimenter the role of the prospective owner. These two roles are not necessarily mutually exclusive—the experimenter may do a considerable portion of the design work. The basic requirements and primary objectives of the experiment are formulated by the experimenter; the experimenter may or may not be aware of the possible alternative approaches that can be followed in the conduct of his experiment. It is the designer's function to make the experimenter aware of these alternatives and to indicate the potential advantages and disadvantages of each of the alternative approaches. It is, however, the experimenter's task to reach the final decision about the conduct of the experiment.

The individual best qualified to design an experiment is the one who is (1) most familiar with the nature of the experimental material, (2) most familiar with the possible alternative methods for designing the experiment, (3) most capable of evaluating the potential advantages and disadvantages of the alternatives. Where an individual possesses all these qualifications, the roles of experimenter and designer are one. On some research problems in many experimental fields, the experimenter is capable of making all the necessary decisions without seeking extensive assistance. On more complex research problems, the experimenter may turn to colleagues who are equally or more familiar with the subject-matter area for assistance in formulating the basic requirements and primary objectives of his experiment. Problems on the design of the experiment may also be discussed with the

subject-matter specialist, and considerable assistance on design problems may be obtained from this source. The experimenter may also turn to the individual whose specialized training is in the area of experimental design, just as the prospective builder turns to the architect for assistance on design problems. If the designer is familiar with the nature of the experimental material and the outcome of past experimentation in the general area of the experiment, he is in a better position to assist the experimenter in evaluating the possible choices as well as to suggest feasible alternative choices.

In the design of experiments there is ample opportunity for ingenuity in the method of attacking the basic problems. Two experiments having identical objectives may be designed in quite different ways: at the same cost in terms of experimental effort, one design may lead to unambiguous results no matter what the outcome, whereas the second design could potentially lead to ambiguous results no matter what the outcome. How good one design is relative to a second for handling the same general objective may be measured (1) in terms of the relative cost of the experimental effort and (2) in terms of the relative precision with which conclusions may be stated. More precise conclusions do not always demand the greater experimental effort, but they generally do demand more careful attention to experimental design.

Without an adequate experimental design, potentially fruitful hypotheses cannot be tested with any acceptable degree of precision. Before rejecting a hypothesis in a research field, one should examine the structure of the experiment to ascertain whether or not the experiment provided a real test of the hypothesis. On the other hand the most carefully planned experiment will not compensate for the lack of a fruitful hypothesis to be tested. In the latter case, the end product of this well-designed experiment can yield only relatively trivial results.

One of the primary objectives of this book is to provide the prospective experimenter with some of the basic principles used in the construction of experimental designs. These principles apply in all areas of experimental work. By the use of these principles, an extensive collection of relatively standard designs have been constructed to handle problems in design that have been encountered in a variety of experiments. These standard designs will be considered in detail, and their potential applications in research areas in the behavioral sciences will be indicated. Seldom does an experimenter have an experiment that is a perfect fit to a standard design. Some modification is frequently required; this is particularly true in experimental work in the area of the behavioral sciences. Careful planning by both the experimenter and the designer is often required in order to cast an experiment in a form that will permit the utilization of a standard design or to modify standard designs in a manner that will more closely meet the requirements of the experiment.

Principles of experimental design have their roots primarily in the logic of

scientific method. Indeed logicians have made substantial contributions to the principles of experimental design. The steps from logic to mathematics are small ones. The now classic work on the basic statistical principles underlying experimental design is R. A. Fisher's *The Design of Experiments*. This work includes more than purely mathematical arguments—it probes into the basic logical structure of experiments and examines the manner in which experiments can provide information about problems put to experimental test. Depending upon how the experiment is conducted, it may or may not provide information about the issues at question. What has become standard working equipment for the individuals specializing in the area of experimental design stems in large measure from this and other works of R. A. Fisher.

What is perhaps the equivalent of a master collection of architect's plans is to be found in the work *Experimental Designs* by W. G. Cochran and G. M. Cox. This work is more than a mere collection of designs. It is a carefully prepared and well-organized text and reference book. Illustrative material is drawn from many different research areas, although most of the material is from the field of agriculture.

The statistical theory underlying major aspects of experimental design is by no means complete. The current literature in the area is extensive.

CHAPTER 1

Basic Concepts in Statistical Inference

1.1 Basic Terminology in Sampling

A statistical population is the collection of all elements about which one seeks information. Only a relatively small fraction, or *sample*, of the total number of elements in a statistical population can generally be observed. From data on the elements that are observed, conclusions or inferences are drawn about the characteristics of the entire population. In order to distinguish between quantities computed from observed data and quantities which characterize the population, the term *statistic* will be used to designate a quantity computed from sample data, and the term *parameter* will be used to designate a quantity characteristic of a population. Statistics are computed from sample data for two purposes: (1) to describe the data obtained in the sample, and (2) to estimate or test hypotheses about characteristics of the population.

If all the elements in a statistical population were measured on a characteristic of interest, and if the measurements were then tabulated in the form of a frequency distribution, the result would be the population distribution for the characteristic measured. A description of the population distribution is made in terms of parameters. The number of parameters necessary to describe the population depends on the form of the frequency distribution. If the form is that of the normal distribution, two parameters will completely describe the frequency distribution—the population mean, designated μ , and the population standard deviation, designated σ . If the form is not normal, the mean and the standard deviation may not be sufficient to describe the distribution. Indeed these two parameters may provide relatively little information about the distribution; other parameters may be required.

The sample mean, designated \bar{X} , generally provides an estimate of the population mean μ . In these same cases, the sample standard deviation, designated s , generally provides an estimate of the population standard deviation σ . The accuracy, or precision, of estimates of this kind depends upon the size of the sample from which such estimates are computed, the

manner in which the sample was drawn from the population, the characteristics of the population from which the sample was drawn, and the formula used to estimate the parameter.

If a sample is drawn in such a way that (1) all elements in the population have an equal and constant chance of being drawn on all draws and (2) all possible samples have an equal (or a fixed and determinable) chance of being drawn, the resulting sample is a *random* sample from the specified population. By no means should a random sample be considered a haphazard, unplanned sample. Numerous other methods exist for drawing samples. Random samples have properties which are particularly important in statistical work. This importance stems from the fact that random sampling ensures constant and independent probabilities; the latter are relatively simple to handle mathematically.

Suppose that one were to draw a large number of samples (say, 100,000), each having n elements, from a specified population. Suppose further that the procedures by which the samples are drawn are comparable for all samples. For each of the samples drawn, suppose that the sample mean \bar{X} and the sample variance s^2 are computed. The frequency distribution of the \bar{X} 's defines operationally what is meant by the *sampling distribution* of the sample mean. A distribution constructed in this way provides an *empirically determined* sampling distribution for the mean. The frequency distribution of the sample variances would provide an empirically determined sampling distribution for the variance. The sampling distribution of a statistic depends, in part, upon the way in which the samples are drawn.

Sampling distributions of statistics are generally tabulated in terms of cumulative frequencies, relative frequencies, or probabilities. The characteristics of sampling distributions are also described by parameters. Frequently the parameters of sampling distributions are related to the parameters of the population from which the samples are drawn. The mean of the sampling distribution is called the *expected value* of the statistic. The standard deviation of the sampling distribution is called the *standard error* of the statistic. The form of the sampling distribution as well as the magnitude of its parameters depends upon (1) the distribution of the measurements in the basic population from which the sample was drawn, (2) the sampling plan followed in drawing the samples, and (3) the number of elements in the sample.

Suppose that the basic population from which sample elements are drawn can be considered to be approximately normal in form, with mean equal to some value μ , and with standard deviation equal to some value σ . In other words, the frequency distribution of the measurements of interest is approximately normal in form, with specified values for the parameters. A normal distribution having a mean equal to μ and a standard deviation equal to σ is designated by $N(\mu, \sigma)$. If one were to draw a large number of random samples of size n from a population in which the measurements

have the approximate form $N(\mu, \sigma)$, the sampling distribution of the statistic \bar{X} would be approximately normal in form, with expected value approximately equal to μ , and with standard error approximately equal to σ/\sqrt{n} . Thus the sampling distribution of the mean of random samples of size n from the approximate population $N(\mu, \sigma)$ would be approximately $N(\mu, \sigma/\sqrt{n})$. This result may be verified by empirical sampling experiments.

The sampling distribution of the statistic \bar{X} , assuming random sampling from the exact population $N(\mu, \sigma)$, can be derived mathematically from the properties of random samples; from purely mathematical considerations it can be shown that this sampling distribution is exactly $N(\mu, \sigma/\sqrt{n})$. Herein lies the importance of random samples—they have properties which permit the estimation of sampling distributions from purely mathematical considerations without the necessity for obtaining empirical sampling distributions. Estimates obtained from such samples have highly desirable properties—the latter will be discussed in a later section. Such purely mathematical considerations lead to scientifically useful results only when the experimental procedures adequately conform to the mathematical models used in predicting experimental results. Also, from purely mathematical considerations, it can be shown that the statistic $(n-1)s^2/\sigma^2$ will have a sampling distribution that corresponds to the chi-square distribution which has $n-1$ degrees of freedom. This last prediction may also be verified by sampling experiments.

If the population distribution is only approximately normal in form, the mathematical sampling distributions just discussed provide approximations to their operational counterparts; the larger the sample size, the better the approximation. One of the basic theorems in sampling theory, the *central-limit theorem*, states that the sampling distribution of the means of random samples will be approximately normal in form regardless of the form of the distribution in the population, provided that the sample size is sufficiently large and provided that the population variance is finite. The more the population distribution differs from a bell-shaped distribution, the larger the sample size must be for the theorem to hold.

Statistics obtained from samples drawn by means of sampling plans which are not random have sampling distributions which are either unknown or which can only be approximated with unknown precision. Good approximations to sampling distributions of statistics are required if one is to evaluate the precision of the inferences made from sample data.

1.2 Basic Terminology in Statistical Estimation

Numerical values of parameters can be computed directly from observed data only when measurements on all elements in the population are available. Generally a parameter is estimated from statistics based upon one or more samples. Several criteria are used to evaluate how good a statistic is as an

estimate of a parameter. One such criterion is lack of bias. A statistic is an *unbiased estimate* of a parameter if the expected value of the sampling distribution of the statistic is equal to the parameter of which it is an estimate. Thus the concept of unbiasedness is a property of the sampling distribution and not strictly a property of a single statistic. When one says that a given statistic is an unbiased estimate of a parameter, what one implies is that in the long run the mean of such statistics computed from a large number of samples of equal size will be equal to the parameter.

The mean \bar{X} of a random sample from a normal population is an unbiased estimate of the population mean because the sampling distribution of \bar{X} has an expected value equal to μ . Suppose that a random sample of size n is drawn from a specified normal population; suppose that the mean of this sample is 45. Then 45 is an unbiased estimate of the population mean. Suppose that a second random sample of size n is drawn from the same population; suppose that the mean of the second sample is 55. Then 55 is also an unbiased estimate of the population mean. Thus two random samples provide two unbiased estimates of the population mean; these estimates will not, in general, be equal to one another. There is no way of deciding which one, considered by itself, is the better estimate. The best single estimate of the population mean, given the two samples, is the average of the two sample means. This average is also an unbiased estimate of the population mean. It is a better estimate of μ in the sense that it has greater precision.

The *precision* of an estimator is generally measured by the standard error of its sampling distribution. The smaller the standard error, the greater the precision. Of two unbiased estimators whose sampling distributions have the same form, the better estimator is the one having the smaller standard error. The standard error of a sampling distribution is a good index of the precision only in those cases in which the form of the distribution approaches the normal distribution as the sample size increases. For statistics whose sampling distribution has this property, the *best* unbiased estimator is defined to be the one having the smallest standard error. The *efficiency* of an unbiased estimator is measured relative to the square of the standard error of the best unbiased estimator. For example, if the squared standard error of one unbiased estimator is σ^2/n and the squared standard error of the best unbiased estimator is $\sigma^2/2n$, then the efficiency of the first estimator is defined to be

$$E_f = \frac{\sigma^2/2n}{\sigma^2/n} = \frac{1}{2}.$$

The concept of *consistency* in an estimator is in a sense related to that of unbiasedness. An estimator is a *consistent* estimate of a parameter if the probability that it differs from the parameter by any amount approaches zero as the sample size increases. In other words, a statistic is a consistent

estimator if the bias tends toward zero as the sample size increases. An unbiased estimator is a consistent estimator. On the other hand, a consistent estimator may be biased for small samples.

Properties of estimators which hold as the sample size increases are called *asymptotic* properties. How large the sample size must be before asymptotic properties can be reasonably expected to hold varies as a function of the characteristics of the population and the method of sampling being used. Consistent estimators are asymptotically unbiased estimators. Where the bias of a consistent estimator is low but its precision is high, the consistent statistic may be used in preference to an unbiased estimator having less precision.

A parameter is, in most cases, a number. It may be estimated by a number, called a *point estimate* of the parameter. Another way of estimating a parameter is to specify a range of numbers, or an interval, within which the parameter lies. This latter type of estimate is known as an *interval estimate* of the parameter. The difference between the largest and smallest numbers of the interval estimate defines the range, or width, of the interval. The sampling distribution of a statistic obtained by means of purely mathematical considerations will provide information about the relative frequency (probability) of statistics in a given interval. Probabilities obtained directly from such sampling distributions provide predictions about the relative frequency with which statistics of given magnitudes will occur, assuming that conditions specified in the mathematical derivation are true in the population. Thus knowledge of sampling distributions permits one to argue from a specified population to consequences in a series of samples drawn from this population.

In statistical estimation, the objective is to obtain estimates of the parameters in the population, given the observations in the sample. The parameters are unknown. Given the magnitude of certain statistics computed from the observed data, from which of several possible alternative populations was this sample drawn? Concepts of likelihood, confidence, inverse probability, and fiducial probability are used by some statisticians to evaluate the answer to this last question. This question can be rephrased in terms of two of these concepts.

1. Given a sample, what is the *likelihood* that it was drawn from a population having a specified set of parameters?

2. Given a sample, with what *confidence* can it be said that the population from which it was drawn has a specified parameter within a given range?

The likelihood of obtaining a given sample is the probability of obtaining the sample as a function of different values of the parameters underlying the population. Admittedly there is only a single set of parameters underlying a specified population. These values are, however, unknown. The relative frequency with which certain samples will occur depends upon the true values of these parameters. Under one set of assumptions about

the parameter values, a given sample may have very high probability of occurring, whereas under a second set of assumptions the probability of the occurrence of a given sample may be very low.

R. A. Fisher introduced a widely used principle in statistical estimation: one selects as an estimator of a parameter that value which will maximize the likelihood of the sample that is actually observed to occur. Estimators having this property are known as *maximum-likelihood estimators*. In many areas of statistics, the principle of maximum likelihood provides estimators having maximum precision (i.e., minimum standard error).

An interval estimate is frequently referred to as a *confidence interval* for a parameter. The two extreme points in this interval, the upper and lower confidence bounds, define a range of values within which there is a specified likelihood (or level of confidence) that that parameter will fall. Given information from a single sample, the parameter either does or does not lie within this range. The procedure by which the upper and lower confidence bounds are determined will, in the long run (if the study is repeated many times) ensure that the proportion of correct statements is equal to the level of confidence for the interval. The numerical values of the upper and lower confidence bounds change from sample to sample, since these bounds depend in part upon statistics computed from the samples.

An interval estimate of a parameter provides information about the precision of the estimate; a point estimate does not include such information. The principles underlying interval estimation for a parameter are closely related to the principles underlying tests of statistical hypotheses.

1.3 Basic Terminology in Testing Statistical Hypotheses

A *statistical hypothesis* is a statement about a statistical population which, on the basis of information obtained from observed data, one seeks to support or refute. A *statistical test* is a set of rules whereby a decision about the hypothesis is reached. Associated with the decision rules is some indication of the accuracy of the decisions reached by following the rules. The measure of the accuracy is a probability statement about making the correct decision when various conditions are true in the population in which the hypothesis applies.

The design of an experiment has a great deal to do with the accuracy of the decisions based upon information supplied by an experiment. The decision rules depend in part upon what the experimenter considers critical bounds on arriving at the wrong decision. However, a statistical hypothesis does not become false when it exceeds such critical bounds, nor does the hypothesis become true when it does not exceed such bounds. Decision rules are guides in summarizing the results of a statistical test—following such guides enables the experimenter to attach probability statements to his decisions. In evaluating the outcome of a single experiment or in using the information in a single experiment as a basis for a course of action, whether an outcome

exceeds an arbitrary critical value may or may not be relevant to the issue at hand. Probability statements that are associated with decision rules in a statistical test are predictions as to what may be expected to be the case if the conditions of the experiment were repeated a large number of times.

The logic of tests on statistical hypotheses is as follows: One assumes that the hypothesis that one desires to test is true. Then one examines the consequences of this assumption in terms of a sampling distribution which depends upon the truth of this hypothesis. If, as determined from the sampling distribution, observed data have relatively high probability of occurring, the decision is made that the data do not contradict the hypothesis. On the other hand, if the probability of an observed set of data is relatively low when the hypothesis is true, the decision is that the data tend to contradict the hypothesis. Frequently the hypothesis that is tested is stated in such a way that, when the data tend to contradict it, the experimenter is actually demonstrating what it is that he is trying to establish. In such cases the experimenter is interested in being able to reject or nullify the hypothesis being tested.

The *level of significance* of a statistical test defines the probability level that is to be considered too low to warrant support of the hypothesis being tested. If the probability of the occurrence of observed data (when the hypothesis being tested is true) is smaller than the level of significance, then the data are said to contradict the hypothesis being tested, and a decision is made to reject this hypothesis. Rejection of the hypothesis being tested is equivalent to supporting one of the possible alternative hypotheses which are not contradicted.

The hypothesis being tested will be designated by the symbol H_1 . (In some notation systems this hypothesis has been designated by the symbol H_0 .) The set of hypotheses that remain tenable when H_1 is rejected will be called the alternative hypothesis and will be designated by the symbol H_2 . The decision rules in a statistical test are with respect to the rejection or non-rejection of H_1 . The rejection of H_1 may be regarded as a decision to accept H_2 ; the nonrejection of H_1 may be regarded as a decision against the acceptance of H_2 . If the decision rules reject H_1 when in fact H_1 is true, the rules lead to an erroneous decision. The probability of making this kind of error is at most equal to the level of significance of the test. Thus the level of significance sets an upper bound on the probability of making a decision to reject H_1 when in fact H_1 is true. This kind of erroneous decision is known as a *type 1 error*; the probability of making a type 1 error is controlled by the level of significance.

If the decision rules do not reject H_1 , when in fact one of the alternative hypotheses is true, the rules also lead to an erroneous decision. This kind of error is known as a *type 2 error*. The potential magnitude of a type 2 error depends in part upon the level of significance and in part upon which one of the possible alternative hypotheses actually is true. Associated with

each of the possible alternative hypotheses is a type 2 error of a different magnitude. The magnitude of a type 1 error is designated by the symbol α , and the magnitude of the type 2 error for a specified alternative hypothesis is designated by the symbol β . The definitions of type 1 and type 2 errors may be summarized as follows:

Decision	State of affairs in the population	
	H_1 true	H_1 false H_2 true
Reject H_1 Accept H_2	Type 1 error (α)	No error
Do not reject H_1 Do not accept H_2	No error	Type 2 error (β)

In this summary, rejection of H_1 is regarded as being equivalent to accepting H_2 and nonrejection of H_1 equivalent to not accepting H_2 . The possibility of a type 1 error exists only when the decision is to reject H_1 ; the possibility of a type 2 error exists only when the decision is not to reject H_1 .

The experimenter has the level of significance (type 1 error) directly under his control. Type 2 error is controlled indirectly, primarily through the design of the experiment. If possible, the hypothesis to be tested is stated in such a way that the more costly error is type 1 error. It is desirable to have both types of error small. However, the two types of error are not independent—the smaller numerically the type 1 error, the larger numerically the potential type 2 error.

To see the relationship between the two types of error, consider Fig. 1.1. In part *a* of this figure the left-hand curve represents the sampling distribution of a relevant statistic when H_1 is true, and the right-hand curve represents the sampling distribution of the same statistic when a particular H_2 is true. The region of rejection of H_1 is defined with reference to the sampling distribution which assumes that H_1 is true. The decision rules specify that H_1 is to be rejected if an observed statistic has any value in the region of rejection. The probability of a statistic's falling in this region is equal to α when H_1 is true. The type 2 error associated with the particular H_2 represented in part *a* is numerically equal to the area under the right-hand curve which falls in the region of nonrejection of H_1 .

In part *b* the numerical value of α is smaller than that in part *a*. This means that the decision rule has smaller type 1 error. The area under the right-hand curve in part *b* that falls in the region of nonrejection of H_1 is larger than the corresponding area in part *a*. Decreasing the numerical value of the type 1 error (level of significance) will increase the potential magnitude of the type 2 error.

The *power* of a test with respect to a specified alternative hypothesis is numerically equal to 1 minus the probability of a type 2 error. Represented geometrically, the power of a test is the area of the sampling distribution, when H_2 is true, that falls in the region of rejection of H_1 . In part *a* of the figure this is the area under the right-hand curve that is to the right of the critical value. The power of a test decreases as the numerical value of α decreases.

The power of a test may be defined symbolically as

$$\text{Power} = P(\text{decision rejects } H_1 \mid H_2 \text{ true}).$$

In words, power is the probability that the decision rule rejects H_1 when a specified H_2 is true. Each of the possible hypotheses in H_2 has its own

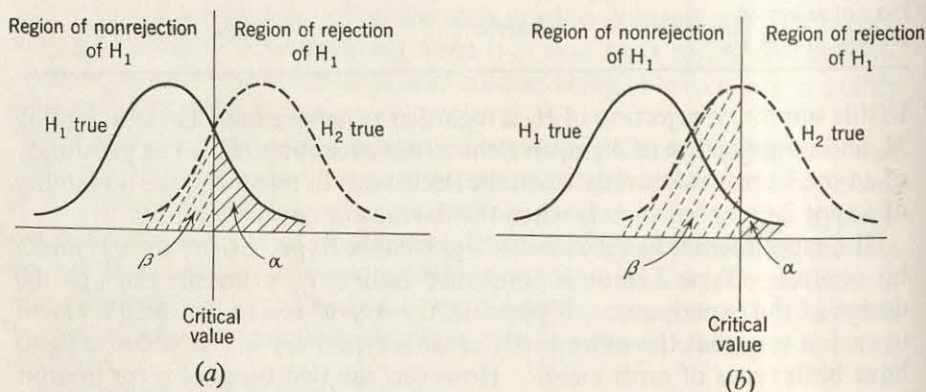


Figure 1.1

power. The closer an alternative hypothesis is to H_1 , that is, the greater the overlap of the corresponding sampling distributions, the lower will be the power of the test with respect to that alternative. A well-designed experiment will have relatively high power with respect to all alternatives which are different in a practical sense from H_1 . For example, if H_1 states that there is zero difference between two means, then one of the possible alternative hypotheses is that the difference is .001 unit. For all practical purposes this alternative may not be different from H_1 ; hence power with respect to this alternative need not be of concern to the experimenter. However, an alternative hypothesis which states that the difference is 5 units may have practically important consequences if true. Power with respect to this alternative would be a matter of concern to the experimenter.

In research in the area of the behavioral sciences, it is often difficult to evaluate the relative costs of type 1 and type 2 in terms of meaningful units. Both kinds of errors may be equally important, particularly in exploratory work. Too much emphasis has been placed upon the level of significance of a test and far too little emphasis upon the power of the test. In many

cases where H_1 is not rejected, were the power of such tests studied carefully, the decisions might more appropriately have been that the experiment did not really provide an adequately sensitive (powerful) test of the hypothesis.

No absolute standards can be set up for determining the appropriate level of significance and power that a test should have. The level of significance used in making statistical tests should be gauged in part by the power of practically important alternative hypotheses at varying levels of significance. If experiments were conducted in the best of all possible worlds, the design of the experiment would provide adequate power for any predetermined level of significance that the experimenter were to set. However, experiments are conducted under the conditions that exist within the world in which one lives. What is needed to attain the demands of the well-designed experiment may not be realized. The experimenter must be satisfied with the best design feasible within the restrictions imposed by the working conditions. The frequent use of the .05 and .01 levels of significance is a matter of a convention having little scientific or logical basis. When the power of tests is likely to be low under these levels of significance, and when type 1 and type 2 errors are of approximately equal importance, the .30 and .20 levels of significance may be more appropriate than the .05 and .01 levels.

The evidence provided by a single experiment with respect to the truth or falsity of a statistical hypothesis is seldom complete enough to arrive at a decision which is free of all possible error. The potential risks in decisions based upon experimental evidence may in most cases be evaluated. What the magnitude of the risks should be before one takes a specified action in each case will depend upon existing conditions. The data from the statistical test will provide likelihoods associated with various actions.

CHAPTER 2

Testing Hypotheses about Means and Variances

2.1 Testing Hypotheses on Means— σ Assumed Known

To illustrate the basic procedures for making a statistical test, a highly simplified example will be used. Suppose that experience has shown that the form of the distribution of measurements on a characteristic of interest in a specified population is approximately normal. Further suppose, given data on a random sample of size 25 from this population, that information about the population mean μ is desired. In particular the experimenter is interested in finding out whether or not the data support the hypothesis that μ is greater than 50.

The first step in the test is to formulate H_1 and H_2 . Suppose that an erroneous decision to reject the hypothesis that the population mean is 50 is more costly than an erroneous decision to reject the hypothesis that the mean is greater than 50. In this case H_1 is chosen to be $\mu = 50$; this choice for H_1 makes the more costly type of error the type I error, which is under the direct control of the experimenter. The alternative hypothesis in this case is $\mu > 50$. The decision rules for this test are to be formulated in such a way that rejection of H_1 is to provide evidence in favor of the tenability of H_2 .

The choice for H_1 could also be $\mu \leq 50$. However, if the data tend to reject the hypothesis that $\mu = 50$ and support the hypothesis that $\mu > 50$, then the data will also tend to reject the hypothesis that $\mu < 50$. Thus, in formulating a decision rule which rejects H_1 only when the data support the hypothesis that $\mu > 50$, only the hypothesis that $\mu = 50$ need be considered. In essence the case $\mu < 50$ is irrelevant (inadmissible) in formulating the decision rule. However, nonrejection of H_1 would imply $\mu \leq 50$.

When it is true that $\mu = 50$, the sampling distribution of the mean of random samples from a normal population is normal in form, with expected value equal to 50 and standard error equal to the population standard deviation divided by the square root of the sample size. In practice the

value of the population standard deviation will not be known, but to keep this example simple, suppose that the population standard deviation σ is equal to 10. Then the standard error of the sampling distribution of the mean for samples of size 25 is $\sigma/\sqrt{n} = 10/\sqrt{25} = 2$. Decision rules must now be formulated to indicate when observed data are consistent with H_1 and when observed data are not consistent with H_1 . The decision rules must indicate a range of potentially observable values of \bar{X} for which the decision will be to reject H_1 . This range of values of \bar{X} will be called the region of rejection for H_1 . The probability of observing an \bar{X} in this region is to be at most equal to the level of significance of the test, i.e., the magnitude of the type 1 error. This sets an upper bound on the probability of reaching

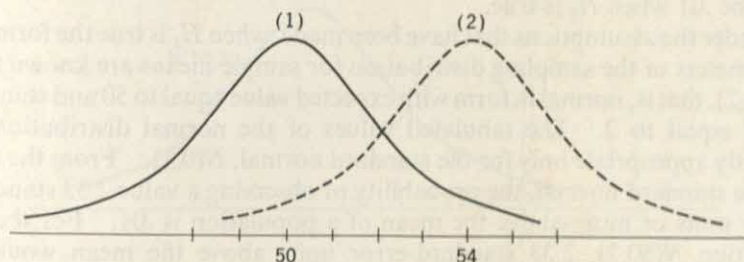


Figure 2.1

the wrong decision when H_1 is true. In addition to satisfying this condition with respect to type 1 error, the region of rejection for H_1 must have relatively high probability for the observed \bar{X} when H_2 is true. Hence the decision rules must specify a range of values of potentially observable \bar{X} in which (1) the probability of an observed \bar{X} 's falling in this region is at most equal to the level of significance when H_1 is true and (2) the probability of an \bar{X} 's falling in this region is relatively high when H_2 is true. The latter condition is necessary to assure the power of the test.

Probabilities associated with the sampling distribution of \bar{X} when H_1 is true are required in order to construct the decision rules. In addition, some knowledge about the relative location of the sampling distribution of \bar{X} when each of the possible alternative hypotheses is true is required. Consider Fig. 2.1. When H_1 is true, the sampling distribution of \bar{X} is given by (1). When H_2 is true (that is, μ is greater than 50), the sampling distribution of \bar{X} will have an expected value somewhere to the right of 50. In particular, one possibility for this expected value is that $\mu = 54$. This possibility is represented by (2). Areas under these curves represent probabilities. The probability of observing an \bar{X} in a range of values covered by the extreme right-hand tail of (1) is relatively low when H_1 is true but relatively higher when the alternative hypothesis is true.

Suppose that the experimenter wants to formulate a set of decision rules which, in the long run, will make the probability of an erroneous decision

when H_1 is true less than .01. This is another way of saying that the level of significance of the test is to be .01. Suppose that the mean of the potentially observable sample is designated by the symbol \bar{X}_{obs} . Then the decision rules will take the following form:

Reject H_1 when \bar{X}_{obs} is greater than L .

Do not reject H_1 otherwise.

L is the critical value for \bar{X}_{obs} ; L must have the property that

$$P(\bar{X}_{\text{obs}} > L \mid H_1 \text{ true}) = .01.$$

In words, the probability of drawing a sample whose mean is greater than L is to be .01 when H_1 is true.

Under the assumptions that have been made, when H_1 is true the form and parameters of the sampling distribution for sample means are known to be $N(50, 2)$, that is, normal in form with expected value equal to 50 and standard error equal to 2. The tabulated values of the normal distribution are directly appropriate only for the standard normal, $N(0, 1)$. From the table of the standard normal, the probability of observing a value 2.33 standard-error units or more above the mean of a population is .01. For the distribution $N(50, 2)$, 2.33 standard-error units above the mean would be $50 + 2.33(2) = 54.66$. Therefore,

$$P(\bar{X} > 54.66 \mid H_1 \text{ true}) = .01.$$

Thus the region of rejection for H_1 is $\bar{X} > 54.66$. When H_1 is true, the probability that a random sample of size 25 from $N(50, 10)$ will have a mean larger than 54.66 is less than .01. When one of the alternative hypotheses is true, i.e., when μ is greater than 50, the probability of a sample mean falling in this region will be higher than .01; the larger the difference between the true value of μ and 50, the higher the probability of an observed mean falling in the region of rejection.

The steps in the formulation of the decision rule have been as follows:

1. Basic population of measurements assumed to be normal in form, with $\sigma = 25$

2. Random sample of size $n = 25$ elements to be drawn from this population.

3. \bar{X}_{obs} to be computed from sample data.

The hypothesis being tested, the alternative hypothesis, the level of significance of the test, and the decision rules are as follows:

$$H_1: \mu = 50.$$

$$H_2: \mu > 50.$$

$$\alpha = .01.$$

Decision rules:

Reject H_1 when $\bar{X}_{\text{obs}} > 54.66$.

Do not reject H_1 otherwise.

The region of rejection for H_1 may be represented geometrically as the right-hand tail of the sampling distribution for \bar{X} which assumes H_1 to be true (see Fig. 2.2).

There are many regions in which the probability is .01 for observing a sample mean. The level of significance of a test does not determine where the region of rejection is to be located. The choice of the extreme right-hand tail of the sampling distribution which assumes H_1 to be true was necessary in order to minimize type 2 error (or, equivalently, to maximize the power). In general, the alternative hypothesis determines the *location* of the region of rejection, whereas the level of significance determines the *size* of the region of rejection. In this case the alternative hypothesis does not

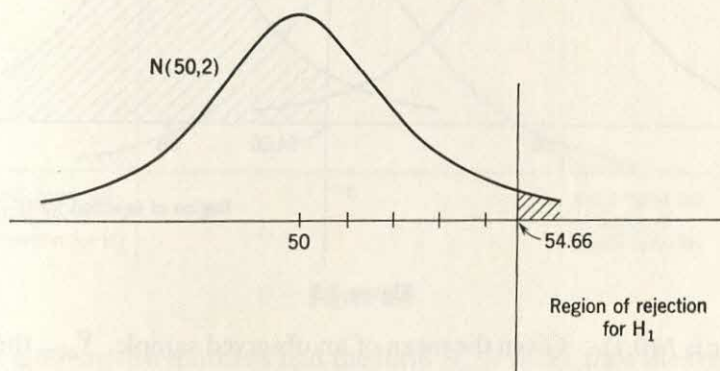


Figure 2.2

include the possibility that μ is less than 50. No matter how much smaller than 54.66 the observed sample mean is, H_1 is not rejected. Thus, if H_1 is not rejected, the evidence would indicate that μ is equal to or less than 50. On the other hand, if H_1 is rejected, the evidence would indicate that μ is greater than 50. Locating the region of rejection for H_1 in the right-hand tail provides maximum power with respect to the alternative hypothesis that μ is greater than 50.

The power of these decision rules with respect to various alternative hypotheses is readily computed. For example, the power with respect to the alternative hypothesis $\mu = 58$ is represented geometrically by the shaded area under curve (2) in Fig. 2.3. This area represents the probability of an observed mean's being greater than 54.66 when the true sampling distribution is $N(58,2)$. With reference to the latter sampling distribution, the point 54.66, which determines the region of rejection, is $(54.66 - 58.00)/2$ or 1.67 standard-error units below the mean. The area from the mean to 1.67 standard-error units below the mean is .45. Hence the total shaded area is $.45 + .50 = .95$. Thus the power of this test with respect to the alternative hypothesis $\mu = 58$ is .95. Conversely, the probability of a type 2 error when $\mu = 58$ is .05.

Instead of working directly with the sampling distribution of the statistic \bar{X} and formulating the decision rules in terms of the statistic \bar{X} , it is more convenient to work with the statistic

$$z = \frac{\bar{X} - \mu_1}{\sigma/\sqrt{n}},$$

where μ_1 is the value specified by H_1 . The sampling distribution of this z

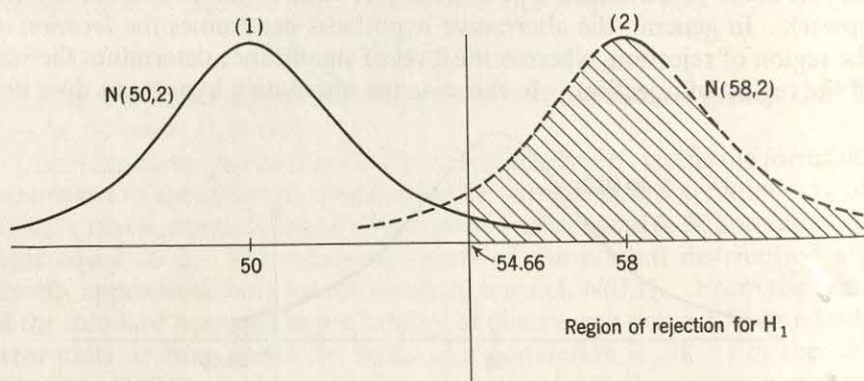


Figure 2.3

statistic is $N(0,1)$. Given the mean of an observed sample, \bar{X}_{obs} , the corresponding value of the z statistic, when H_1 is $\mu = 50$ and $\sigma/\sqrt{n} = 2$, is

$$z_{\text{obs}} = \frac{\bar{X}_{\text{obs}} - 50}{2}.$$

If the alternative hypothesis is $H_2: \mu > 50$, then the decision rules for a test having level of significance .01 are as follows:

Reject H_1 when $z_{\text{obs}} > 2.33$.

Do not reject H_1 otherwise.

The value 2.33 satisfies the condition that

$$P(z_{\text{obs}} > 2.33 \mid H_1 \text{ true}) = .01.$$

This numerical value actually is the 99th centile point on $N(0,1)$ and will be designated by the symbol $z_{.99}$. Thus $z_{.99} = 2.33$. Since the level of significance for this test is $\alpha = .01$, the critical value for the decision rule can be designated $z_{1-\alpha}$, which in this case is $z_{.99}$.

For the general case in which the region of rejection for H_1 is the right-hand tail of $N(0,1)$ and the level of significance is equal to some value α , the decision rules take the following form:

Reject H_1 when $z_{\text{obs}} > z_{1-\alpha}$.

Do not reject H_1 otherwise.

Suppose that the mean for the sample observed actually is 60. Then the numerical value of the z statistic (when H_1 is that $\mu = 50$ and $\sigma/\sqrt{n} = 2$) is

$$z_{\text{obs}} = \frac{60 - 50}{2} = 5.00.$$

Since z_{obs} is larger than 2.33, H_1 is rejected. Hence the observed data do not support the hypothesis that the population mean is 50. The data indicate that the mean in the population is greater than 50.

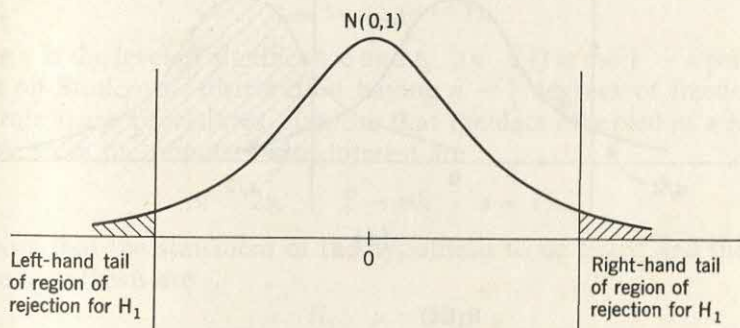


Figure 2.4

If the alternative hypothesis had the form $H_2: \mu \neq 50$, then the region of rejection for H_1 would have the form

$$z_{\text{obs}} < z_{\alpha/2} \quad \text{and} \quad z_{\text{obs}} > z_{1-(\alpha/2)}.$$

For this kind of alternative hypothesis, the region of rejection for H_1 includes both the left-hand and right-hand extreme tails of the sampling distribution associated with H_1 . The two parts of the region of rejection for H_1 are sketched in Fig. 2.4. For example, if $\alpha = .01$, the two-tailed region of rejection for H_1 would be z_{obs} less than and $z_{.005}$ and z_{obs} greater than $z_{.995}$. Locating the region of rejection for H_1 in this manner provides power with respect to the possibility that μ is less than 50, as well as to the possibility that μ is greater than 50. An alternative hypothesis of this form is called a two-tailed alternative hypothesis, and tests which admit to the possibility of a two-tailed alternative hypothesis are called two-tailed tests. The size of either tail of the region of rejection is equal to one-half the level of significance; the total size of the region of rejection is equal to the level of significance.

In cases in which the experimenter is interested in rejecting H_1 only when the alternative hypothesis is one having a specified direction with respect to H_1 , a one-tailed rather than a two-tailed alternative hypothesis is the more appropriate. Limiting the region of rejection to one tail of the sampling distribution for H_1 provides greater power with respect to an

alternative hypothesis in the direction of that tail. This fact is illustrated geometrically in Fig. 2.5. The power under a two-tailed test with respect to a specified alternative hypothesis to the right of zero is shown by the shaded area in part *a*. The corresponding power with respect to a one-tailed test is shown in part *b*. Although the magnitude of the type 1 error is the same in both cases, the increased power in the one-tailed case is at the expense of zero power with respect to alternative hypotheses which are to

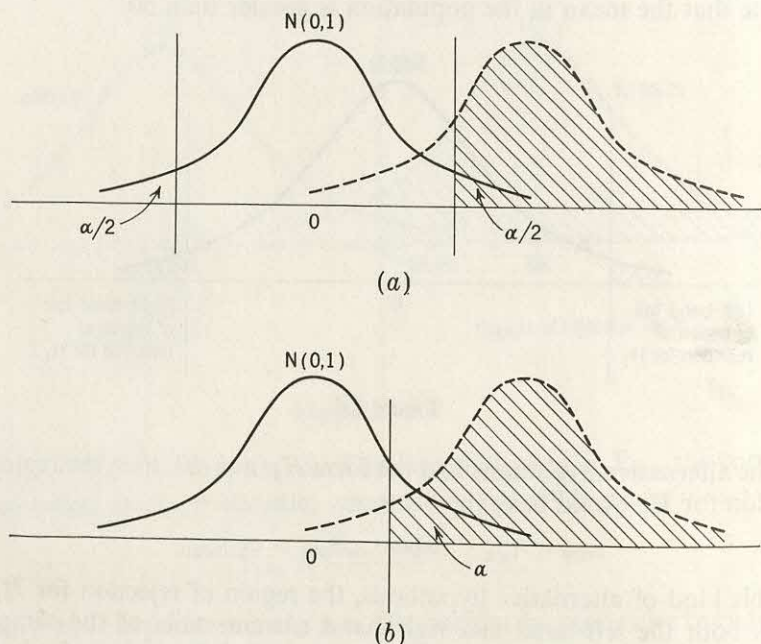


Figure 2.5

the left of zero. In the latter case, all hypotheses corresponding to sampling distributions to the left of zero may be considered part of H_1 .

2.2 Tests of Hypotheses on Means— σ Estimated from Sample Data

In Sec. 2.1, the value of the population standard deviation σ was assumed to be known. Generally σ is estimated by the sample standard deviation s . Testing the hypothesis considered in the last section when σ is estimated by s involves no basic change in principles for arriving at a decision rule. The sampling distribution of the statistic that must be used, however, is no longer normal in form. The test statistic convenient for use in this case is

$$t = \frac{\bar{X} - \mu_1}{s/\sqrt{n}},$$

where μ_1 is the value of μ specified by H_1 . When H_1 is true, the sampling distribution of this t statistic is Student's t distribution having $n - 1$ degrees of freedom. (The degrees of freedom for this sampling distribution are determined by the degrees of freedom for s .) When H_1 is not true, the form of the sampling distribution for this t statistic is no longer approximated by Student's t distribution but rather by what is called a *noncentral t distribution*.

For the one-tailed alternative hypothesis $\mu > \mu_1$, the region of rejection for H_1 is given by

$$t_{\text{obs}} > t_{1-\alpha}(n-1),$$

where α is the level of significance and $t_{1-\alpha}(n-1)$ is the $1 - \alpha$ percentile point on Student's t distribution having $n - 1$ degrees of freedom. To illustrate these procedures, suppose that the data observed in a random sample from the population of interest are

$$n = 25, \quad \bar{X} = 60, \quad s = 15.$$

Suppose that the statement of the hypothesis to be tested and the alternative hypothesis are

$$H_1: \mu = 50$$

$$H_2: \mu > 50$$

and that the level of significance of the test is .01. From the table of the distribution of the t statistic having 24 degrees of freedom one finds that

$$P(t_{\text{obs}} > 2.49 \mid H_1 \text{ true}) = .01.$$

That is, the table of the t distribution indicates that $t_{.99}(24) = 2.49$. Hence the decision rules for this test are as follows:

Reject H_1 when t_{obs} is larger than 2.49.

Do not reject H_1 otherwise.

From the sample data, t_{obs} is found to be

$$t_{\text{obs}} = \frac{60 - 50}{15/\sqrt{25}} = 3.33.$$

Since t_{obs} is greater than the critical value 2.49, t_{obs} falls in the region of rejection for H_1 . Hence the decision rules indicate that H_1 should be rejected.

The interpretation of this test is as follows: On the basis of the data in a random sample of size 25 from a population of interest, the hypothesis that $\mu = 50$ cannot be considered tenable when the test is made at the .01 level of significance. If this hypothesis were true, the probability of obtaining the data in the sample would be less than .01. The data obtained support the hypothesis that the mean of the population is greater than 50.

Having rejected the hypothesis that $\mu = 50$, suppose that the experimenter wanted to find the largest value for μ_1 which would lead to non-rejection of H_1 . The region of nonrejection for H_1 is defined by the inequality

$$\frac{\bar{X} - \mu_1}{s/\sqrt{n}} \leq t_{1-\alpha}(n-1).$$

Solving this inequality for μ_1 gives

$$\mu_1 \geq \bar{X} - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1).$$

Thus any value of μ_1 equal to or greater than $\bar{X} - (s/\sqrt{n})t_{1-\alpha}(n-1)$ will yield a t statistic that will fall in the region of nonrejection for H_1 . For the numerical example just considered, any H_1 that specifies μ_1 to be equal to or greater than

$$60 - (3.00)(2.49) = 52.53$$

would, on the basis of the single sample observed, lead to a decision not to reject H_1 . Thus, on the evidence supplied by the single sample observed, any value for μ_1 equal to or greater than 52.53 would make t_{obs} smaller than the critical value of 2.49. Therefore the experimenter may conclude that the population mean is likely to be greater than 52.53. If the experimenter were to test hypotheses specifying that μ_1 is any value equal to or less than 52.53, the decision in every case (for the data in the given sample) would be to reject H_1 . This conclusion may be expressed in the form of a one-tailed confidence interval on the population mean. This confidence interval takes the general form

$$C\left[\mu \geq \bar{X} - \frac{s}{\sqrt{n}} t_{1-\alpha}(n-1)\right] = 1 - \alpha.$$

The numerical values in terms of the observed sample data and $\alpha = .01$ are

$$C[\mu \geq 52.53] = .99.$$

The value 52.53 may be considered as the lower bound for μ . If one were to draw additional samples, the mathematical form of the lower bound would remain the same but its numerical value would change, since the numerical values of \bar{X} and s would change. Once a sample has been drawn and numerical values for the confidence interval determined, the statement made in the confidence interval is either true or false. However, the procedure by which the confidence interval is constructed will, in the long run, lead to statements which are correct with probability equal to $1 - \alpha$.

The example that has just been considered involved a one-tailed alternative hypothesis. Suppose that the experimenter is willing to reject H_1 when μ is either smaller or larger than 50. In this case the alternative

hypothesis takes the form $\mu \neq 50$. To provide power with respect to both tails of the alternative hypothesis, the decision rules for this case are as follows:

$$\text{Reject } H_1 \text{ when } t_{\text{obs}} \begin{cases} < t_{\alpha/2}(n-1). \\ > t_{1-(\alpha/2)}(n-1). \end{cases}$$

Otherwise do not reject H_1 .

The region of rejection for H_1 for the case of a two-tailed alternative hypothesis is sketched in Fig. 2.6. The size of the region of rejection in each tail is equal to $\alpha/2$. The left-hand tail of the region of rejection makes provision for power with respect to alternative hypotheses $\mu < 50$; the right-

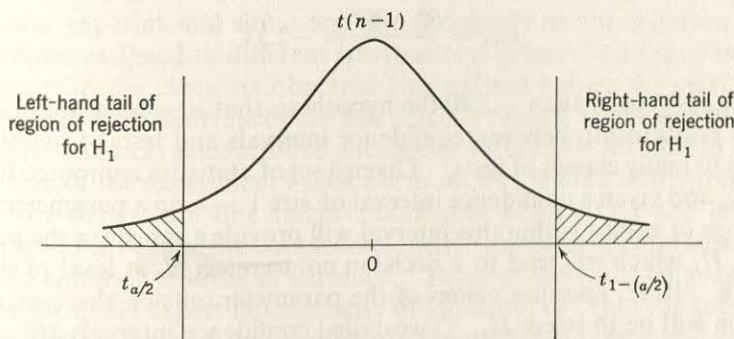


Figure 2.6

hand tail of the region of rejection makes provision for power with respect to alternative hypotheses $\mu > 50$.

For $n - 1 = 24$ and $\alpha = .01$,

$$t_{\alpha/2} = t_{.005} = -2.80,$$

and

$$t_{1-(\alpha/2)} = t_{.995} = 2.80.$$

(Since Student's t distribution is symmetrical, $t_{\alpha/2} = -t_{1-(\alpha/2)}$.) For this case the decision rules are

$$\text{Reject } H_1 \text{ when } t_{\text{obs}} \begin{cases} < -2.80. \\ > 2.80. \end{cases}$$

Otherwise do not reject H_1 .

For $\bar{X}_{\text{obs}} = 60$ and $s_{\text{obs}} = 15$, $t_{\text{obs}} = 3.33$. Hence the decision rules lead to rejecting H_1 . Having rejected the hypothesis that $\mu = 50$, the experimenter may be interested in determining the range of values of μ which, on the basis of the observed sample data, would not be rejected by these

decision rules. This range of values is defined by a two-tailed confidence interval on μ , which is given by

$$C[\bar{X} - c \leq \mu \leq \bar{X} + c] = 1 - \alpha,$$

where

$$c = \frac{s}{\sqrt{n}} t_{1-(\alpha/2)}(n-1).$$

The numerical value of this confidence interval for the observed sample and $\alpha = .01$ is

$$C[51.60 \leq \mu \leq 68.40] = .99.$$

To illustrate the fact that any value of μ in this range which is selected for H_1 leads to a decision not to reject H_1 , consider the hypothesis that $\mu = 68$. For this H_1

$$t_{\text{obs}} = \frac{60 - 68}{3} = -2.33.$$

Since t_{obs} is greater than -2.80 the hypothesis that $\mu = 68$ is not rejected.

This relationship between confidence intervals and tests of hypotheses applies to many classes of tests. Given a set of statistics computed from a sample, and given a confidence interval of size $1 - \alpha$ on a parameter, then the range of values within this interval will provide a range for the parameter in H_1 which will lead to a decision not to reject H_1 at level of significance α . If H_1 specifies values of the parameter outside this range, the decision will be to reject H_1 . Two-tailed confidence intervals are associated with two-tailed tests and one-tailed confidence intervals with one-tailed tests. Thus confidence intervals provide information about the potential outcomes of a series of individual tests.

The power of tests associated with Student's distribution is more difficult to compute than the power of tests associated with the normal distribution. Since the sampling distribution of the t statistic when H_1 is not true is a noncentral t distribution, computation of the power with respect to alternative hypotheses requires tables of the noncentral t distribution. Illustrations of the applications of this latter distribution are given in Johnson and Welch (1940).

2.3 Testing Hypotheses about the Difference between Two Means—Assuming Homogeneity of Variance

One problem common to many fields of research may be cast in the following form: Which one of two procedures will produce the better results when used in a specified population? To provide information relevant for an answer, the experimenter may draw two samples from the specified population. His experiment might consist of following procedure A in one of the samples and procedure B in the other sample. (These procedures will be referred to as treatments A and B .) The question,

which one of the two is the better, requires some criterion on which to base the answer. Several criteria may be relevant—treatment A may be better with respect to some of these criteria, treatment B better with respect to others. Techniques exist for the evaluation of several criteria simultaneously, but in this section methods for evaluating a single criterion at a time will be considered.

Suppose that the experimenter measures each of the elements in the two samples on a single criterion. The results may be that some of the scores under treatment A are higher than those under B , and vice versa. If the distribution of the scores within each of the samples is approximately normal in form, then a comparison of the means of the criterion scores provides one kind of information about which of the two treatments gives the better results.

In some experimental situations, the variability on the criterion within the samples assigned to different treatments is primarily a function of (1) differences in the elements observed that existed before the start of the experiment—such differences are not directly related to the experimental treatment—and (2) uncontrolled sources of variability introduced during the course of the experiment which are in no way related to the treatment itself. In such cases one might reasonably expect that the criterion variance within each of the samples assigned to the experimental treatments is due to common sources of variance. In more technical language, one might reasonably expect homogeneity of variance, i.e., that the sources of variance within each of the samples are essentially the same and that the variances in the corresponding populations are equal.

It is convenient to formalize the arguments just given in more mathematical terms. The formal mathematical argument will serve to make explicit what it is that one assumes to arrive at the conclusions that have just been reached. Let the criterion measure on element i in the sample given experimental treatment j be designated by the symbol X_{ij} . In this case there are two experimental treatments; so j stands for either treatment A or treatment B . Suppose that this measurement may be expressed as the sum of a quantity τ_j , which represents the effect of experimental treatment j , and a quantity ε_{ij} , which is not directly related to experimental treatment j . That is, suppose that

$$(1) \quad X_{ij} = \tau_j + \varepsilon_{ij}.$$

The effect ε_{ij} includes all the unique characteristics associated with the element i as well as all uncontrolled effects associated with the experimental conditions under which the measurement is made. The effect τ_j is assumed to be constant for all elements in the experimental group assigned to treatment j , whereas the effect ε_{ij} varies from element to element within the group and is in no direct way related to the experimental treatment. The effect ε_{ij} is frequently called the experimental error.

The term X_{ij} on the left-hand side of (1) represents a quantity that is observed. The terms on the right-hand side of (1) cannot be observed—they designate the variables that account for what is observed. The terms τ_j and ε_{ij} represent structural variables underlying the observed data; (1) is referred to as a structural model. The first basic assumption that has been made about the variables in the structural model is that they are uncorrelated. (Being uncorrelated is a less stringent assumption than statistical independence. The latter assumption is, however, required in making tests. For purposes of estimation, only the assumption of zero correlation is required.)

Let σ_a^2 and σ_b^2 designate the expected values of the criterion variance within the respective experimental groups. (That is, σ_a^2 represents the mean of the sampling distribution of the statistic s_a^2 , the variance on the criterion for a sample of elements given treatment A .) Suppose that the experiment is designed in such a way that in the long run there will be no difference in the unique characteristics of the group of elements assigned to treatments A and B . Suppose also that the experiment is conducted in such a manner that the uncontrolled sources of variability are comparable in the two experimental groups. These latter assumptions imply homogeneity of experimental error; i.e., they imply that $\sigma_a^2 = \sigma_b^2$. These latter assumptions also imply that the expected value of the mean of the experimental error within treatment groups A and B will be equal. That is, if $\bar{\varepsilon}_a$ and $\bar{\varepsilon}_b$ represent the respective sample means of the experimental error within the treatment groups, the expected values of these quantities will be equal if the assumptions are true. The quantities $\bar{\varepsilon}_a$ and $\bar{\varepsilon}_b$ are the means of structural variables and cannot be computed directly from a single sample.

Experience has shown that the model (1) and the assumptions made about the variables in the model are appropriate for a large class of experimental situations. The tests to be considered in this section are suitable for this class. In terms of this structural model, the mean of the criterion scores for a sample of elements given treatment A may be represented as

$$(2) \quad \bar{X}_a = \tau_a + \bar{\varepsilon}_a.$$

Since the effect τ_a is assumed to be constant for all elements in the sample, its mean effect will be simply τ_a . The corresponding mean for a sample of elements given treatment B may be represented by

$$(3) \quad \bar{X}_b = \tau_b + \bar{\varepsilon}_b.$$

The difference between the two sample means has the form

$$(4) \quad \bar{X}_a - \bar{X}_b = (\tau_a - \tau_b) + (\bar{\varepsilon}_a - \bar{\varepsilon}_b).$$

In words, (4) says that the observed difference between the criterion mean for the sample given treatment A and the criterion mean for the sample given treatment B is in part a function of the difference in effectiveness of

the two treatments and in part a function of the difference between the average experimental error associated with each of the means. One purpose of a statistical test in this context is to find out whether the observed difference is of a magnitude that may be considered a function of experimental error alone or whether the observed difference indicates some effect larger than that due to experimental error.

The details in the analysis of an experiment designed to study the effects of treatments A and B on a specified criterion will now be considered. From the population of interest a random sample of n elements is drawn. The elements are subdivided at random into two subsamples—one of size n_a , the other of size n_b . (In most cases n_a will be equal to n_b . The most sensitive design makes n_a and n_b proportional to σ_a^2 and σ_b^2 , respectively.) The sample of size n_a is assigned to experimental treatment A ; the sample of size n_b is assigned to experimental treatment B . After the administration of the treatments, each of the elements in the experiment is measured on a common criterion. Suppose that experience in related research indicates that the distribution of such criterion measures tends to be approximately normal in the population of interest. Suppose that inspection of the observed data does not contradict what past experimentation indicates about the form of the distribution of the criterion measures within each of the experimental conditions.

To summarize the information from the experiment, the mean and standard deviation for each of the experimental groups are computed. The sample statistics \bar{X}_a and s_a^2 provide, respectively, estimates of the parameters μ_a and σ_a^2 . Similarly for treatment group B , \bar{X}_b and s_b^2 provide, respectively, estimates of the parameters μ_b and σ_b^2 . A schematic outline of this experiment is given in Table 2.3-1. If the form of the distribution of the measurements were not approximately bell-shaped, other statistics might be more appropriate to summarize the information in the samples.

If one were to repeat this experiment a large number of times, each time starting with a different random sample, and if one were to compute the statistic $\bar{X}_a - \bar{X}_b$ for each experiment, the sampling distribution of this statistic would be found to be approximately normal in form. The expected value of the sampling distribution would be approximately $\mu_a - \mu_b$, and the standard error would be $\sqrt{(\sigma_a^2/n_a) + (\sigma_b^2/n_b)}$. Since σ_a^2 is assumed to be equal to σ_b^2 , let the common value of the population variance be designated by σ_e^2 . In terms of σ_e^2 , the standard error of the sampling distribution would have the form

$$(5) \quad \sigma_{\bar{X}_a - \bar{X}_b} = \sqrt{\sigma_e^2 \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}.$$

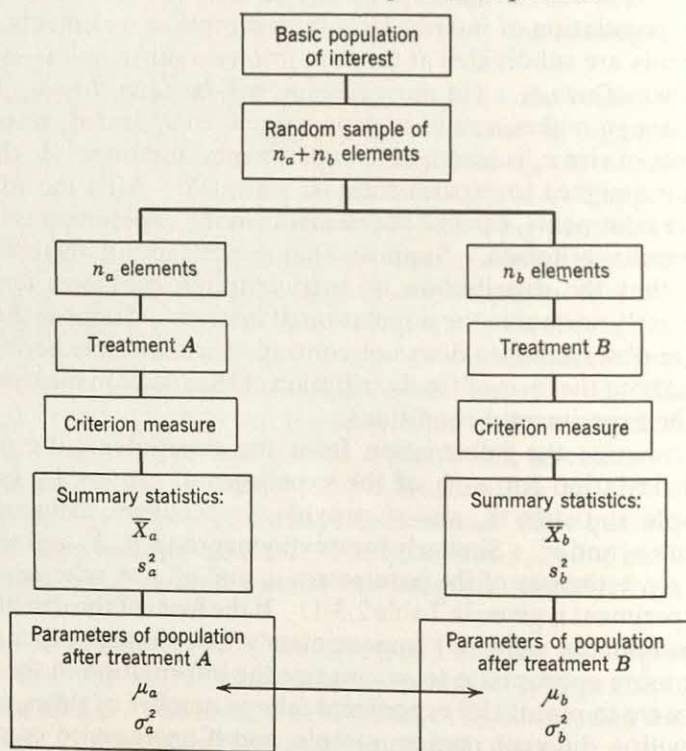
The value of the common population variance σ_e^2 is not known. Given s_a^2 and s_b^2 computed from sample data, the best estimate of σ_e^2 is a weighted average of the sample variances, the weights being the respective degrees of

freedom. Designating this weighted average by the symbol s_p^2 ,

$$(6) \quad s_p^2 = \frac{(n_a - 1)s_a^2 + (n_b - 1)s_b^2}{(n_a - 1) + (n_b - 1)}.$$

This weighted average of the sample variances is known as the pooled within-class estimate of the common population variance. The degrees of freedom

Table 2.3-1 Outline of Steps in an Experiment



for s_p^2 are the sum of the respective degrees of freedom for the parts; i.e.,

$$(n_a - 1) + (n_b - 1) = n_a + n_b - 2.$$

The statistic used to test hypotheses about $\mu_a - \mu_b$ is

$$(7) \quad t = \frac{(\bar{X}_a - \bar{X}_b) - (\mu_a - \mu_b)}{\sqrt{s_p^2(1/n_a + 1/n_b)}}.$$

The sampling distribution of this statistic, under the assumptions that have been made, is the t distribution having $n_a + n_b - 2$ degrees of freedom. Operationally, if this experiment were to be repeated a large number of times, each time with a different random sample from a specified population, and if one actually knew the value of $\mu_a - \mu_b$, then the distribution of the

resulting t statistics could be approximated by the t distribution having $n_a + n_b - 2$ degrees of freedom. The latter degrees of freedom are those associated with s_p^2 . In general one does not know the value of $\mu_a - \mu_b$. Its value must be specified by H_1 .

The denominator of the t statistic defined in (7) is often symbolized by $s_{\bar{x}_a - \bar{x}_b}$, that is,

$$s_{\bar{x}_a - \bar{x}_b} = \sqrt{s_p^2 \left(\frac{1}{n_a} + \frac{1}{n_b} \right)}.$$

Under this notation system,

$$t = \frac{(\bar{X}_a - \bar{X}_b) - (\mu_a - \mu_b)}{s_{\bar{x}_a - \bar{x}_b}}.$$

The t distribution is used to test hypotheses about $\mu_a - \mu_b$. In terms of the right-hand side of the structural model in (4), the expected value of numerator of the t statistic in (7) is $E(\bar{\varepsilon}_a - \bar{\varepsilon}_b) = 0$, and the expected value of the denominator is $\sigma_{\bar{\varepsilon}_a - \bar{\varepsilon}_b}$. Hence, when the hypothesized value of $\mu_a - \mu_b$ is actually the true value of the difference between these parameters, the t statistic provides a standardized measure of the difference in the average experimental error for the two experimental conditions.

Suppose that the experimenter is interested in testing the following hypothesis:

$$H_1: \mu_a - \mu_b = \delta.$$

$$H_2: \mu_a - \mu_b \neq \delta.$$

$$\text{Level of significance} = \alpha.$$

The numerical value of δ is the smallest practically important difference of interest to the experimenter. (This value is often taken to be zero.) Since the alternative hypothesis is two-tailed, a two-tailed region of rejection for H_1 is required. The region of rejection for H_1 is defined by the two tails of the sampling distribution which assumes H_1 to be true. The decision rules are as follows.

$$\text{Reject } H_1 \text{ when } t_{\text{obs}} \begin{cases} < t_{\alpha/2}(n_a + n_b - 2). \\ > t_{1-(\alpha/2)}(n_a + n_b - 2). \end{cases}$$

Do not reject H_1 otherwise.

The t statistic will be numerically large when (1) H_1 is not true or (2) H_1 is true but the difference between the mean experimental errors is unusually large relative to what is expected on the basis of the assumptions underlying the experimental design. The probability of rejecting H_1 when the latter contingency occurs is less than the level of significance of the test. The region of rejection for H_1 is sketched in Fig. 2.7.

To illustrate the use of the t statistic and its sampling distribution in making the test about the difference $\mu_a - \mu_b$, suppose that the following

data are obtained from an experiment designed and conducted under conditions such that the assumptions underlying the sampling distribution of the t statistic are satisfied:

Treatment	Sample size	Sample variance	Criterion mean
A	$n_a = 8$	$s_a^2 = 18$	$\bar{X}_a = 20$
B	$n_b = 10$	$s_b^2 = 12$	$\bar{X}_b = 25$

The experimenter is interested in making a two-tailed test on the hypothesis that $\mu_a - \mu_b = 0$. Since rejecting this hypothesis when it is true is considered by the experimenter to be a more costly error than not rejecting this hypothesis when it is false, H_1 has the form $\mu_a - \mu_b \neq 0$. The level of significance for this test is chosen to be .05. The pooled estimate of the

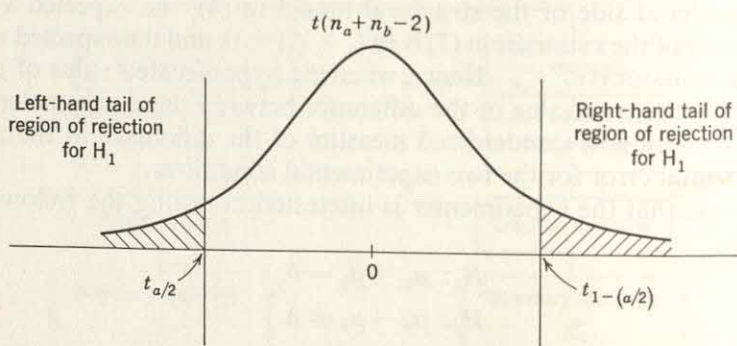


Figure 2.7

common population variance will have 16 degrees of freedom. Hence the decision rules for this test will have the following form:

$$\text{Reject } H_1 \text{ when } t_{\text{obs}} \begin{cases} < -2.12. \\ > 2.12. \end{cases}$$

Do not reject H_1 otherwise.

The value 2.12 is the .975 percentile point on the t distribution having 16 degrees of freedom, that is, $t_{.975}(16) = 2.12$. Since the t distribution is symmetrical, $t_{.025}(16) = -2.12$.

For these data, using (6), the estimate of the population variance is

$$s_p^2 = \frac{7(18) + 9(12)}{16} = 14.62.$$

The value of t_{obs} is given by

$$t_{\text{obs}} = \frac{(20 - 25) - 0}{\sqrt{14.62(\frac{1}{8} + \frac{1}{10})}} = \frac{-5}{\sqrt{14.62(\frac{18}{80})}} = \frac{-5}{1.82} = -2.75.$$

Since t_{obs} is less than the critical value -2.12 , H_1 is rejected. Thus the hypothesis that treatments A and B are equally effective with respect to the criterion measure is not supported by the experimental data. Inspection indicates that treatment B has the higher mean. Hence the data support the alternative hypothesis that $\mu_a - \mu_b < 0$, that is, $\mu_a < \mu_b$.

The hypothesis that the difference between the population means is zero having been rejected, the tenable values for this difference are given by a confidence interval for $\mu_a - \mu_b$. This interval has the general form

$$C[(\bar{X}_a - \bar{X}_b) - c \leq \mu_a - \mu_b \leq (\bar{X}_a - \bar{X}_b) + c] = 1 - \alpha,$$

where

$$c = t_{1-(\alpha/2)} S_{\bar{X}_a - \bar{X}_b}.$$

For the numerical data being considered,

$$c = 2.12(1.82) = 3.86.$$

Hence a .95 confidence interval on the difference between the two treatment means is

$$C[-8.86 \leq \mu_a - \mu_b \leq -1.14] = .95.$$

For these sample data, any H_1 which specifies that $\mu_a - \mu_b$ is within this interval will lead to nonrejection of H_1 when $\alpha = .05$.

2.4 Computational Formulas for the t Statistic

For simplicity in computational work, the basic formula for the t statistic is not the most suitable for use. To avoid rounding error at intermediate stages of work, the best computational procedure puts off divisions and square roots until the final stages. For accurate computational work the general rule is to do all possible addition, subtraction, and multiplication before doing any division or taking any square root. To arrive at a computational formula for the t statistic, the following notation will be used:

ΣX_a = sum of the n_a observations in experimental group A .

ΣX_b = sum of the n_b observations in experimental group B .

$\Sigma X = \Sigma X_a + \Sigma X_b$ = sum of all observations in the experiment.

ΣX_a^2 = sum of the squares of the observations in experimental group A .

ΣX_b^2 = sum of the squares of the observations in experimental group B .

$\Sigma X^2 = \Sigma X_a^2 + \Sigma X_b^2$ = sum of the squares of all observations in the experiment.

The definition of the symbol L_a is

$$(1) \quad L_a = n_a \Sigma X_a^2 - (\Sigma X_a)^2.$$

An analogous definition of the symbol L_b is

$$(2) \quad L_b = n_b \Sigma X_b^2 - (\Sigma X_b)^2.$$

In terms of this notation, a computational formula for the square of the t statistic, when H_1 has the form $\mu_a - \mu_b = 0$, is

$$(3) \quad t^2 = \frac{(n_a + n_b - 2)(n_b \Sigma X_a - n_a \Sigma X_b)^2}{(n_a + n_b)(n_b L_a + n_a L_b)}.$$

Taking the square root of t^2 yields the t statistic. The algebraic sign is given by the sign of $\bar{X}_a - \bar{X}_b$.

In terms of L_a and L_b computational formulas for variances are

$$(4) \quad s_a^2 = \frac{L_a}{n_a(n_a - 1)},$$

$$(5) \quad s_b^2 = \frac{L_b}{n_b(n_b - 1)}.$$

A computational formula for s_p^2 , which minimizes rounding error at intermediate stages, is

$$(6) \quad s_p^2 = \frac{n_a n_b \Sigma X^2 - n_b (\Sigma X_a)^2 - n_a (\Sigma X_b)^2}{n_a n_b (n_a + n_b - 2)}.$$

There is one disadvantage to these formulas: the numbers involved at intermediate steps tend to become quite large even for relatively small samples. Use of these formulas is illustrated in the numerical example given in Table 2.4-1.

Table 2.4-1 Numerical Example of Computation of t Statistic

Treatment A		Treatment B	
	3		6
	5		5
$n_a = 7$	2	7	$n_b = 10$
	4	8	
	6	9	
	2	4	
	7	7	
		8	
		9	
		7	
	$\Sigma X_a = 29$	$\Sigma X_b = 70$	
	$\Sigma X_a^2 = 143$	$\Sigma X_b^2 = 514$	
$L_a = 7(143) - (29)^2 = 160$		$L_b = 10(514) - (70)^2 = 240$	
$s_a^2 = \frac{160}{7(6)}$	= 3.81	$s_b^2 = \frac{240}{10(9)}$	= 2.67
$\bar{X}_a = \frac{29}{7}$	= 4.14	$\bar{X}_b = \frac{70}{10}$	= 7.00

$$\begin{aligned}
 t^2 &= \frac{(n_a + n_b - 2)(n_b \Sigma X_a - n_a \Sigma X_b)^2}{(n_a + n_b)(n_b L_a + n_a L_b)} \\
 &= \frac{(15)[10(29) - 7(70)]^2}{(17)[10(160) + 7(240)]} = \frac{600,000}{55,760} = 10.760
 \end{aligned}$$

$$\begin{aligned}
 F_{.95}(1, 15) &= 4.54 \\
 t &= -\sqrt{10.760} = -3.28 \dagger \quad t_{.025}(15) = -2.13
 \end{aligned}$$

† Negative sign is used because $\bar{X}_a - \bar{X}_b$ is negative.

The decision rules for $H_1: \mu_a - \mu_b = 0$ against a two-tailed alternative hypothesis may be stated in terms of either t_{obs} or t_{obs}^2 . In terms of t_{obs} the decision rules are (assuming that $\alpha = .05$)

$$\text{Reject } H_1 \text{ if } t_{\text{obs}} \begin{cases} < -2.13. \\ > 2.13. \end{cases}$$

Otherwise do not reject H_1 .

Since $t_{\text{obs}} = -3.28$, H_1 is rejected. Inspection of the data indicates that treatment B gives the greater mean criterion score. In terms of t_{obs}^2 , the decision rules are

$$\text{Reject } H_1 \text{ if } t_{\text{obs}}^2 > 4.54.$$

Otherwise do not reject H_1 .

In this case $t_{\text{obs}}^2 = 10.76$; hence H_1 is rejected. The critical value for the t^2 statistic is $F_{1-\alpha}(1, n_a + n_b - 2)$ or equivalently $t_{1-(\alpha/2)}^2(n_a + n_b - 2)$.

2.5 Test for Homogeneity of Variance

The test on population means developed in Sec. 2.3 was based upon a structural model which assumed that $\sigma_a^2 = \sigma_b^2$. In the absence of extensive information from past experimentation in an area, the data obtained in the experiment are sometimes used to make preliminary tests on the model. Preliminary tests on structural models do not establish the appropriateness of the models; rather their appropriateness depends upon the design of the experiment and the nature of the sources of variation. The purpose of preliminary tests is to provide a partial check on whether or not the observed data tend to be consistent with the model. The observed data may actually be consistent with several models.

Moderate departures from the hypothesis that $\sigma_a^2 = \sigma_b^2$ do not seriously affect the accuracy of the decisions reached by means of the t test given in Sec. 2.3. In more technical language, the t test is *robust* with respect to moderate departures from the hypothesis of homogeneity of variance. An extensive investigation of the effect of unequal variances upon the t test and the corresponding F test is found in the work of Box (1954). The term *moderate* in this context is relative to the magnitude and difference in sample sizes. To illustrate the effect of unequal population variances upon the accuracy of the decision rule based upon the t test, which assumes equal population variances, consider the case in which $\sigma_a^2 = 2\sigma_b^2$. If $n_a = 5$ and $n_b = 5$, then the 95th percentile point on the F distribution which assumes the population variances equal is approximately equal to the 94th percentile point on the sampling distribution which actually takes into account the difference in the population variances. Thus for an .05-level test with these sample sizes, this violation of the assumption that $\sigma_a^2 = \sigma_b^2$ results in an error in the level of significance of approximately 1 per cent (in the direction of rejecting H_1 more often than should be the case).

The work of Box (1954) also indicates that the t test is robust with respect to the assumption of normality of the distributions within the treatment populations. That is, the type 1 error of the decision rule is not seriously affected when the population distributions deviate from normality. Even when population distributions are markedly skewed, the sampling distribution of the t statistic, which assumes normality, provides a good approximation to the exact sampling distribution which takes into account the skewness. In summary, preliminary tests on the structural model for the t test, which

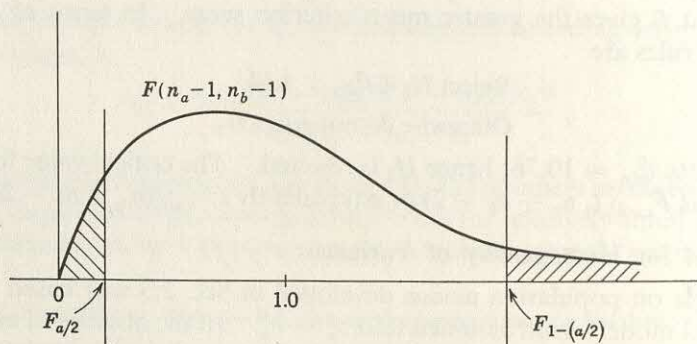


Figure 2.8

assumes homogeneity of variance and normality of distributions, are not of primary importance with respect to type 1 error, particularly preliminary tests for normality of distribution.

The test of the hypothesis that $\sigma_a^2 = \sigma_b^2$ is useful in its own right, quite apart from its use in a preliminary test. Suppose that the following data are potentially available for random samples given treatment A and treatment B :

Treatment	Sample size	Sample variance
A	n_a	s_a^2
B	n_b	s_b^2

Assuming that the distribution of the measurements in the treatment populations is approximately normal, under the hypothesis that $\sigma_a^2 = \sigma_b^2$ the statistic

$$F = \frac{s_a^2}{s_b^2}$$

has a sampling distribution which is approximated by $F(n_a - 1, n_b - 1)$, that is, an F distribution with $n_a - 1$ degrees of freedom for the numerator and $n_b - 1$ degrees of freedom for the denominator. Operationally, if the experiment from which the potential data given above were to be repeated a large number of times, and if for each experiment an F statistic were computed, then the resulting distribution of the F statistics would be approximately $F(n_a - 1, n_b - 1)$. The region of rejection for the hypothesis $\sigma_a^2 = \sigma_b^2$

against the alternative hypothesis $\sigma_a^2 \neq \sigma_b^2$ is sketched in Fig. 2.8. The F distribution is not symmetrical, but there is a relationship by means of which percentile points in the left tail may be obtained from points in the right tail. This relationship is

$$(1) \quad F_{\alpha/2}(n_a - 1, n_b - 1) = \frac{1}{F_{1-(\alpha/2)}(n_b - 1, n_a - 1)}$$

Most tables of the F distribution provide values only for the right-hand tail. Values for the left-hand tail are readily computed by means of the relation (1). There is, however, a procedure by which use of the left-hand tail may be avoided. It is actually immaterial which treatment is called A and which B . Suppose that F_{obs} is defined to be the ratio of the larger sample variance to the smaller sample variance. Hence F_{obs} will always tend to fall toward the right-hand tail of the F distribution for which the degrees of freedom for the numerator are equal to the degrees of freedom for the larger sample variance and the degrees of freedom for the denominator are equal to the degrees of freedom for the smaller sample variance.

To illustrate this test, suppose that the following data were obtained in an experiment:

Treatment	Sample size	Sample variance
A	$n_a = 10$	$s_a^2 = 15.58$
B	$n_b = 8$	$s_b^2 = 28.20$

To test the hypothesis $\sigma_a^2 = \sigma_b^2$ against the alternative hypothesis $\sigma_a^2 \neq \sigma_b^2$,

$$F_{\text{obs}} = \frac{28.20}{15.58} = 1.81.$$

If the level of significance for this test is chosen to be $\alpha = .10$, then the critical value for the region of rejection is $F_{.95}(7,9) = 3.29$. Since F_{obs} does not exceed this value, the hypothesis that $\sigma_a^2 = \sigma_b^2$ is not contradicted by the observed data.

If one draws a large number of random samples of size n_a from an approximately normal population $N(\mu_a, \sigma_a^2)$, then the sampling distribution of the statistic

$$\chi_a^2 = \frac{(n_a - 1)s_a^2}{\sigma_a^2}$$

is approximated by the chi-square distribution having $n_a - 1$ degrees of freedom. If one were to draw similar samples from the population which is approximately $N(\mu_b, \sigma_b^2)$, then the sampling distribution of the statistic

$$\chi_b^2 = \frac{(n_b - 1)s_b^2}{\sigma_b^2}$$

is approximated by the chi-square statistic having $n_b - 1$ degrees of freedom. Assuming that the two populations have no elements in common, the two

chi-square statistics will be independent. In an actual experiment there will be one sample from the A population and one sample from the B population.

The statistic

$$(2) \quad F = \frac{\chi_a^2/(n_a - 1)}{\chi_b^2/(n_b - 1)} = \frac{s_a^2/s_b^2}{\sigma_a^2/\sigma_b^2}$$

has a sampling distribution given by $F(n_a - 1, n_b - 1)$. This statistic may be used to test any hypothesis about the relationship between σ_a^2 and σ_b^2 . Under the hypothesis that $\sigma_a^2 = \sigma_b^2$, the ratio $\sigma_a^2/\sigma_b^2 = 1.00$, and (2) assumes the form $F = s_a^2/s_b^2$. The ratio of two statistics whose sampling distributions are independent chi squares will always have a sampling distribution given by an F distribution.

2.6 Testing Hypotheses about the Difference between Two Means— Assuming That Population Variances Are Not Equal

Suppose that independent random samples of size n_a and n_b are drawn from a common population and are assigned, respectively, to treatment conditions A and B . The elements are then measured on a common criterion. The following data may be obtained from this type of experiment:

Treatment	Sample size	Sample mean	Sample variance
A	n_a	\bar{X}_a	s_a^2
B	n_b	\bar{X}_b	s_b^2

The sample data for treatment A provide estimates of the parameters of a population A , a hypothetical population which consists of the collection of elements after the administration of treatment A . Thus \bar{X}_a is an estimate of μ_a , and s_a^2 is an estimate of σ_a^2 . If the distribution of the criterion measures is approximately normal in form, the population A may be designated as approximately $N(\mu_a, \sigma_a^2)$. The data from sample B are assumed to provide estimates of the parameters of a population which is approximately $N(\mu_b, \sigma_b^2)$. Thus populations A and B are potential populations that could arise from the basic common population from which the samples were drawn.

In Sec. 2.3, tests about $\mu_a - \mu_b$ were made under the assumption that $\sigma_a^2 = \sigma_b^2$. In this section, tests about $\mu_a - \mu_b$ make no assumption about the equality of the population variances. The sampling distribution of the statistic

$$(1) \quad z = \frac{(\bar{X}_a - \bar{X}_b) - (\mu_a - \mu_b)}{\sqrt{(\sigma_a^2/n_a) + (\sigma_b^2/n_b)}}$$

is $N(0, 1)$. Two sets of parameters are involved in this statistic. Hypotheses about $\mu_a - \mu_b$ cannot be tested unless values for σ_a^2 and σ_b^2 are known. If

n_a and n_b are both large (say, larger than 30), then s_a^2 and s_b^2 may be substituted for the corresponding parameters and the resulting statistic has a sampling distribution approximated by $N(0,1)$. For small samples, however, the statistic

$$(1a) \quad t' = \frac{(\bar{X}_a - \bar{X}_b) - (\mu_a - \mu_b)}{\sqrt{(s_a^2/n_a) + (s_b^2/n_b)}}$$

has a sampling distribution which is neither the normal distribution nor Student's t distribution.

Several workers have attempted to obtain the exact sampling distribution of the t' statistic. The end products have differed slightly because of differing assumptions that were made in the course of the derivations. Behrens made one attempt to obtain the sampling distribution of the t' statistic; Fisher enlarged upon the work of Behrens, and the resulting distribution is called the Behrens-Fisher distribution. Tables of the Behrens-Fisher distribution are given in Fisher and Yates (1953, p. 52). A method of approximating critical values of the Behrens-Fisher distribution by Student's t distribution is given by Cochran and Cox (1957, p. 101).

Not all mathematical statisticians agree with the logic underlying the derivation of the Behrens-Fisher distribution. Using a somewhat different approach to a related problem, Satterthwaite (1946) derived an approximation for the t' distribution which involves Student's t distribution having degrees of freedom approximated by the quantity

$$(2) \quad f = \frac{U^2}{[V^2/(n_a - 1)] + [W^2/(n_b - 1)]},$$

where $V = \frac{s_a^2}{n_a}$, $W = \frac{s_b^2}{n_b}$, and $U = V + W$.

The critical values for the t' statistic are obtained from tables of Student's t distribution having degrees of freedom equal to the nearest integer to f . A slightly different (and perhaps closer) approximation for f is given by

$$(3) \quad f = \frac{U^2}{[V^2/(n_a + 1)] + [W^2/(n_b + 1)]} - 2.$$

Welch (1947) has derived an exact sampling distribution for the t' statistic. Tables for this distribution have been prepared by Aspin (1949). Welch has shown that his distribution for the t' statistic may also be approximated by Student's t distribution, the appropriate t distribution having degrees of freedom

$$(4) \quad f = \frac{f_a f_b}{f_b c^2 + f_a (1 - c)^2},$$

where $f_a = n_a - 1$, $f_b = n_b - 1$, and $c = \frac{s_a^2/n_a}{(s_a^2/n_a) + (s_b^2/n_b)}$.

The value of f in (4) will be less than $f_a + f_b$ but greater than the smaller of f_a and f_b . If H_1 is rejected using the degrees of freedom equal to the smaller of f_a and f_b , the value of f need not be computed, since the decision reached by its use will be to reject H_1 . The Welch approximation to the sampling distribution of the t' statistic is from many points of view the soundest. Its use is illustrated in Table 2.6-1.

Table 2.6-1 Numerical Example

	Treatments	
	A	B
Sample size	$n_a = 16$	$n_b = 8$
Sample mean	$\bar{X}_a = 30.00$	$\bar{X}_b = 21.00$
Sample variance	$s_a^2 = 32.00$	$s_b^2 = 80.00$
$H_1: \mu_a - \mu_b = 0$		
$H_2: \mu_a - \mu_b \neq 0$		
$\alpha = .05$		
$t'_{\text{obs}} = \frac{(30.00 - 21.00) - 0}{\sqrt{(32.00/16) + (80.00/8)}} = \frac{9}{\sqrt{12}} = 2.60$		
$c = \frac{32.00/16}{(32.00/16) + (80.00/8)} = .167 \quad f = \frac{(15)(7)}{7(.167)^2 + 15(.833)^2} = 9.90$		

To test the hypothesis that $\mu_a - \mu_b = 0$ against a two-tailed alternative hypothesis, the decision rules for a test with level of significance equal to .05 are as follows:

$$\text{Reject } H_1 \text{ if } t'_{\text{obs}} \begin{cases} < t_{.025}(f) \\ > t_{.975}(f) \end{cases}$$

Otherwise do not reject H_1 .

The nearest integer to the numerical value of f computed in Table 2.6-1 is 10. From tables of Student's t distribution, $t_{.975}(10) = 2.23$. The value of $t'_{\text{obs}} = 2.60$ is greater than 2.23; hence the hypothesis that $\mu_a - \mu_b = 0$ is rejected. In this case it is not actually necessary to compute the value of f . The degrees of freedom for the variance associated with the smaller sample is 7; hence the numerical value of f will be larger than 7. Using $t_{.975}(7) = 2.36$ as a critical value leads to the decision to reject H_1 . This critical value is larger than that based upon f . Hence H_1 will also be rejected when the proper number of degrees of freedom is used in determining the critical value.

The critical value obtained from tables of the Behrens-Fisher distribution is 2.20. The use of the Cochran and Cox approximation to the Behrens-Fisher test will also be illustrated for the data in Table 2.6-1. The critical

value for t'_{obs} is given by

$$(5) \quad t_{\text{critical}} = \frac{w_a t_a + w_b t_b}{w_a + w_b},$$

where $w_a = s_a^2/n_a$, $w_b = s_b^2/n_b$, $t_a = t_{1-(\alpha/2)}(n_a - 1)$, and $t_b = t_{1-(\alpha/2)}(n_b - 1)$. For the data in Table 2.6-6 when $\alpha = .05$, $w_a = 2.00$, $w_b = 10.00$, $t_a = 2.13$, $t_b = 2.36$, and

$$t_{\text{critical}} = \frac{2.00(2.13) + 10.00(2.36)}{12.00} = 2.32.$$

In general t_{critical} computed in this manner will differ only slightly from the critical value obtained by means of the Welch approximation.

Formula (3) is actually a modified version of (4). Using (3) to compute f for the numerical data in Table 2.6-1 gives

$$f = \frac{(12.00)^2}{[(2.00)^2/17] + [(10.00)^2/9]} - 2 = \frac{144}{11.35} - 2 = 10.69.$$

Use of formula (4) gave 9.90. Use of (4) is to be preferred to (3).

2.7 Testing Hypotheses about the Difference between Two Means—Correlated Observations

In Secs. 2.3 and 2.6 the data from two independent samples were used as the basic data for tests. Elements in one of the samples were independent of (in no way related to) elements in the other sample. Hence the data in the two samples were uncorrelated. If the elements in single sample of size n are observed under both treatments A and B , then the resulting data are generally correlated. In this case two measurements are made on each element; hence the term repeated measurements to describe this type of design. If the order in which the treatments are administered has no effect upon the final outcome, then the difference between the two measures on the same element on a common criterion provides a measure of the relative effectiveness of the treatments. This measure is free of variation due to unique but systematic effects associated with the elements themselves. In this respect, each element serves as its own control.

The following structural model, which describes the sources of variation that underlie an observation, is applicable in a variety of experimental situations,

$$(1) \quad X_{ia} = \tau_a + \pi_i + \varepsilon_{ia},$$

where X_{ia} = criterion measure for element i under treatment A ;

τ_a = assumed magnitude of effect of treatment A ;

π_i = unique systematic effect associated with element i ;

ε_{ia} = uncontrolled sources of variance affecting observation X_{ia} (the experimental error).

Each of the structural variables on the right-hand side of (1) is assumed to be independent of the others. The corresponding structural model for an observation on element i under treatment B has the form

$$(2) \quad X_{ib} = \tau_b + \pi_i + \varepsilon_{ib}.$$

In Sec. 2.3, the structural model for an observation combined the term π_i with the experimental error, since π_i could not be distinguished from experimental error. In experimental designs having repeated measures, this source of variance may be eliminated from experimental error. The difference between two measures on the same criterion made on a randomly selected element i is

$$(3) \quad d_i = X_{ia} - X_{ib} = (\tau_a - \tau_b) + (\varepsilon_{ai} - \varepsilon_{bi}).$$

This difference does not involve the variable π_i . The term $\tau_a - \tau_b$ is assumed to be constant for all elements, whereas the term $\varepsilon_{ai} - \varepsilon_{bi}$ varies from element to element. Hence all variance of the d_i 's is a function of $\varepsilon_{ai} - \varepsilon_{bi}$, which is assumed to be a measure of experimental error. The mean of the differences has the structural form

$$(4) \quad \bar{d} = \bar{X}_a - \bar{X}_b = (\tau_a - \tau_b) + (\bar{\varepsilon}_a - \bar{\varepsilon}_b).$$

Thus \bar{d} consists of the sum of two components. One of the components is a measure of the difference between the treatment effects; the other is a measure of the difference between the average experimental error associated with each of the treatments.

If the experimental error for the observations made under treatment A differs in no systematic way from the experimental error for the observations made under treatment B , then it is reasonable to assume that

$$(5) \quad E(\bar{\varepsilon}_a - \bar{\varepsilon}_b) = 0.$$

Therefore,

$$(6) \quad E(\bar{d}) = \tau_a - \tau_b = \mu_a - \mu_b.$$

In words, \bar{d} provides an unbiased estimate of $\mu_a - \mu_b$. If the distribution of the d 's is approximately normal with variance equal to σ_e^2 , the expected value of the variance of the term $\varepsilon_{ai} - \varepsilon_{bi}$ in the structural model, then the sampling distribution of the statistic

$$(7) \quad t = \frac{\bar{d} - (\mu_a - \mu_b)}{\sqrt{s_d^2/n}}$$

may be approximated by Student's t distribution having $n - 1$ degrees of freedom, where n is the number of differences. In (7), s_d^2 is used as an estimate of σ_e^2 .

The application of this sampling distribution to testing hypotheses about $\mu_a - \mu_b$ will be illustrated by means of the data in Table 2.7-1. These data represent the measures on a common criterion before and after the administration of a treatment. In this type of experiment the question of which

treatment is to be administered first is not relevant. For a two-tailed test at the .05 level of significance of the hypothesis $\mu_a - \mu_b = 0$, the decision rules are as follows:

$$\text{Reject } H_1 \text{ if } t_{\text{obs}} \begin{cases} < t_{.025}(6) = -2.45. \\ > t_{.975}(6) = 2.45. \end{cases}$$

Do not reject H_1 otherwise.

The value of $t_{\text{obs}} = 3.61$ falls in the region of rejection for H_1 . Inspection of the data indicates that the difference in the population means is probably greater than zero.

Table 2.7-1 Numerical Example

Person number	Before treatment	After treatment	Difference
1	3	6	3
2	8	14	6
3	4	8	4
4	6	4	-2
5	9	16	7
6	2	7	5
7	12	19	7
			$\Sigma d = 30 \quad \bar{d} = 4.29$
			$\Sigma d^2 = 188$

$$L_d = n \Sigma d^2 - (\Sigma d)^2$$

$$= 7(188) - (30)^2 = 416$$

$$s_d^2 = \frac{L_d}{n(n-1)} = \frac{416}{7(6)} = 9.90$$

$$\sqrt{s_d^2/n} = \sqrt{9.90/7} = 1.19$$

$$H_1: \mu_a - \mu_b = 0$$

$$H_2: \mu_a - \mu_b \neq 0$$

$$\alpha = .05$$

$$t_{\text{obs}} = \frac{4.29 - 0}{1.19} = 3.61$$

A confidence interval for the difference between the population means takes the following general form,

$$C(\bar{X} - c \leq \mu_a - \mu_b \leq \bar{X} + c) = 1 - \alpha,$$

where

$$c = t_{1-(\alpha/2)}(n-1)\sqrt{s_d^2/n}.$$

The symbol $t_{1-(\alpha/2)}(n-1)$ is to be interpreted as the $[1 - (\alpha/2)]$ -percentile point on the t distribution with $n-1$ degrees of freedom, not as $n-1$ times this value. For the data in Table 2.7-1,

$$c = 2.45(1.19) = 2.92.$$

The confidence interval is

$$C(1.37 \leq \mu_a - \mu_b \leq 7.21) = .95.$$

Hypotheses which specify that the difference in the population means lies in this range will lead to a decision not to reject H_1 when the data in Table 2.7-1 are used to make the test.

If a single element is to be given two different treatments for the purpose of evaluating the difference in the effectiveness, there is the question of which treatment is to be given first. It may make some difference in the results if the treatments are administered in the order A followed by B or B followed by A . That is, the carry-over effect, if any, from the sequence A followed by B , or vice versa, may be different. This kind of effect is known as a treatment by order interaction. To check upon the possibility of this interaction in a repeated-measures design, half the elements in the study may be given one sequence, the other half the reverse sequence. This design will permit the evaluation of order effects as well as treatment by order interactions. Tests associated with the evaluation of interaction effects are discussed in Chap. 7. However, if no interaction of this type exists, the t statistic in (7) may be used to test the difference in treatment effects when the order of administration of the treatments is counterbalanced. In the absence of any information about this interaction, the method of attack in Chap. 7 is to be preferred to the use of the t statistic in (7).

The variance s_d^2 in the denominator of (7) may be expressed in terms of the variances of the individual measures, s_a^2 and s_b^2 . This relationship is

$$(8) \quad \begin{aligned} s_d^2 &= s_a^2 + s_b^2 - 2r_{ab}s_as_b \\ &= \text{var}_a + \text{var}_b - 2\text{cov}_{ab}, \end{aligned}$$

where r_{ab} is the product-moment correlation between the measures, and where cov_{ab} is the covariance between the measures. The following relationship holds:

$$\text{cov}_{ab} = r_{ab}s_as_b.$$

The notations var_a and var_b are sometimes used interchangeably with s_a^2 and s_b^2 .

The computation of s_d^2 by means of this formula is illustrated by the numerical data in Table 2.7-1. A summary of the computational steps for the statistics needed is given in Table 2.7-2. Substituting the numerical values in Table 2.7-2 for the corresponding statistics in (8) gives

$$s_d^2 = 32.62 + 12.90 - 2(.868)(5.71)(3.59) = 9.91.$$

Within rounding error, this is the value obtained by working directly with the difference measures.

One of the primary advantages of repeated measures is the potential reduction in variance due to experimental error. For this design

$$(9) \quad s_d^2 = s_{\bar{X}_a - \bar{X}_b}^2 = \frac{s_a^2}{n} = \frac{s_a^2}{n} + \frac{s_b^2}{n} - \frac{2r_{ab}s_as_b}{n}.$$

The corresponding estimate for the case of uncorrelated observations is

$$(10) \quad s_{\bar{X}_a - \bar{X}_b}^2 = \frac{s_a^2}{n} + \frac{s_b^2}{n}.$$

If the correlation is positive, and it will be positive if the model in (1) holds, then the estimate of the experimental error obtained from a design in which (9) is appropriate will be smaller than that obtained from a design involving uncorrelated observations by a factor of $2r_{ab}s_a s_b$. However, the degrees of freedom for the estimate in (9) are $n - 1$, whereas the degrees of freedom for the estimate in (10) are $2n - 2$. Before a repeated-measures design may be

Table 2.7-2 Numerical Example (Continued)

Before treatment (B)	After treatment (A)	Product of treatments
$\Sigma X_b = 44$	$\Sigma X_a = 74$	$\Sigma X_a X_b = 572$
$\Sigma X_b^2 = 354$	$\Sigma X_a^2 = 978$	$L_{ab} = n \Sigma X_a X_b - (\Sigma X_a)(\Sigma X_b)$
$L_b = n \Sigma X_b^2 - (\Sigma X_b)^2$	$L_a = n \Sigma X_a^2 - (\Sigma X_a)^2$	$= 7(572) - (74)(44)$
$= 7(354) - (44)^2$	$= 7(978) - (74)^2$	$= 748$
$= 542$	$= 1370$	$r_{ab} = \frac{L_{ab}}{\sqrt{L_a L_b}}$
$s_b^2 = \frac{L_b}{n(n-1)}$	$s_a^2 = \frac{L_a}{n(n-1)}$	$= \frac{748}{\sqrt{(1370)(542)}}$
$= 12.90$	$= 32.62$	$= .868$
$s_b = 3.59$	$s_a = 5.71$	

considered more efficient than a design which does not have repeated measures, the decrease in the experimental error must be sufficient to offset the reduction in degrees of freedom.

In areas of research in which the variance associated with the term π_i in the model in (2) is likely to be large, the control on this source of variation provided by a repeated-measures design will greatly increase the sensitivity of the experiment. The smaller the proportion of the total variance in an experiment due to effects which are not related to the treatments per se, the more sensitive the experiment.

2.8 Combining Several Independent Tests on the Same Hypothesis

A series of experiments designed to test the same hypothesis may be conducted at different places or at different times. Suppose that the samples used in each of these experiments are independent; i.e., different random samples are employed in each of the experiments. Suppose that the probabilities of the observed outcomes with respect to the common hypothesis

tested are designated P_1, P_2, \dots, P_k . If the experimenter desires to make an over-all probability statement, the following statistic may be used:

$$(1) \quad \chi^2 = 2\sum u_i, \quad \text{where } u_i = -\ln P_i.$$

Under the hypothesis that the observed probabilities are a random sample from a population of probabilities having a mean of .50, the χ^2 statistic in (1) has a sampling distribution which is approximated by the chi-square distribution having $2k$ degrees of freedom.

Use of this statistic is illustrated by the data in Table 2.8-1. These data

Table 2.8-1 Numerical Example

Experiment	t_{obs}^\dagger	Probability	$-\ln(\text{probability})$
1	.87	.20	1.61
2	.54	.30	1.20
3	1.10	.15	1.90
4	1.50	.07	2.66
5	1.30	.11	2.21
			9.58

$$\chi^2 = 2(9.58) = 19.16$$

$$\chi_{.95}^2(10) = 18.3; \quad \chi_{.975}^2(10) = 20.5$$

† df = 14.

give t statistics obtained in a series of five independent experiments testing a specified hypothesis against a one-tailed alternative hypothesis. The probabilities of obtaining the observed t statistic or one more extreme in the direction of the alternative hypothesis is given in the column headed Probability. The natural logarithm of the probability with the sign changed is given in the next column. The numerical value for the chi-square statistic, 19.16, is twice the sum of this last column. k , the number of experiments, is 5; hence the degrees of freedom for this chi-square statistic are $2(5) = 10$. Since the probability of obtaining a chi-square statistic of 19.16 or larger is less than .05 when the mean probability in the population is .50, this latter hypothesis is rejected (if $\alpha = .05$). Thus the combined evidence from the series of experiments indicates that the common H_1 in each of the experiments should be rejected if inferences are to be made with respect to the combined populations represented by the series of experiments.

There is an alternative approach to combining the outcomes of several experiments involving the use of the t statistic based upon relatively large samples. Under the hypothesis that the mean value for the t statistic in the population is zero, the statistic

$$(2) \quad z = \frac{\sum t_i}{\sqrt{k}}$$

has a sampling distribution which is $N(0,1)$. For the data in Table 2.8-1,

$$z = \frac{.87 + .54 + 1.10 + 1.50 + 1.3}{\sqrt{5}} = 2.37.$$

The probability of obtaining this value of z or one larger when the mean of t in the population is zero is less than .01. Hence the combined evidence from the five experiments indicates that the H_1 common to each of the experiments should be rejected if the scope of the inference is the combined populations and if $\alpha = .05$. The sampling distribution of the statistic in (2) is approximately normal in form only when the t_i 's are approximately normally distributed; hence in practice use of the statistic should be limited to cases in which the degrees of freedom for the t_i are larger than 30.

In some instances one could combine the data from all five experiments into a single larger experiment. This procedure might, however, tend to obscure rather than clarify the interpretation—the former would be the case if differences associated with the unique conditions of the individual experiments were allowed to augment the experimental error. This contingency can be avoided through use of designs which will be discussed in the next chapter.

CHAPTER 3

Design and Analysis of Single-factor Experiments

3.1 Introduction

In testing statistical hypotheses or in setting confidence bounds on estimates, one uses sampling distributions determined by purely mathematical considerations. One postulates a model, imposes various conditions upon this model, and then derives the consequences in terms of sampling distributions which are valid for the mathematical system. To the extent that the model and the conditions imposed upon it approximate an actual experiment, the model can be used as a guide in drawing inferences from the data.

If use is to be made of existing, well-developed mathematical models, experiments must be designed in a way that makes these models appropriate. If an experiment cannot meet the specifications in existing models, the experimenter may be able to develop a model tailored to the specific needs of his experiment. But the problem of the analysis of the resulting data must still be solved. If the mathematics needed to derive sampling distributions having known characteristics is manageable, the specially tailored model leads to inferences of known precision. Otherwise inferences drawn from an experiment of this kind will have unknown precision.

The analysis of data obtained from an experiment is dependent upon its design and the sampling distributions appropriate for the underlying population distributions. The design, in part, determines what the sampling distributions will be. Associated with standard designs are analyses justified by the mathematical models which led to the construction of these designs. Alternative designs are often available for an experiment having stated objectives. Depending upon the specific situations, one design may be more efficient than another for the same amount of experimental effort—more efficient in terms of the power of the resulting tests and the narrowness of the resulting confidence intervals. A major problem in planning an experiment is to find or develop that design which is most efficient per unit of experimental effort in reaching the primary objectives of the experiment.

The most efficient design from a purely mathematical point of view may be so costly in terms of time, money, and effort as to render it unworkable. In general, the smaller the variation due to experimental error, other factors being constant, the more efficient the design. This source of variation may be reduced by introducing various kinds of controls or by increasing the sample size. Both methods of reducing experimental error may be used, but which one provides the greater reduction per unit of cost depends upon features unique to the experimental conditions.

The study of the statistical aspects of experimental design will assist the experimenter in finding the most adequate and feasible model for his experiment. This model must permit the experimenter to reach decisions with respect to all the objectives of his experiment. Whether or not a model adequately represents the experimental situation calls for expert knowledge of the subject-matter field. In appraising the adequacy of alternative models, the experimenter is often made aware of sources of variation and possible implications that he had not thoroughly considered. Thus design problems often force the experimenter to formulate his experiment in terms of variables that will lead to clear-cut interpretations for the end product.

Some of the criteria for good experimental designs are as follows:

1. The analyses resulting from the design should provide unambiguous information on the primary objectives of the experiment. In particular the design should lead to unbiased estimates.
2. The model and its underlying assumptions should be appropriate for the experimental material.
3. The design should provide maximum information with respect to the major objectives of the experiment per minimum amount of experimental effort.
4. The design should provide some information with respect to all the objectives of the experiment.
5. The design must be feasible within the working conditions that exist for the experimenter.

The designs that will be considered in this chapter are appropriate for what have been called single-factor experiments. The primary objective of this kind of experiment is to compare the relative effectiveness of two or more treatments on a common criterion. The term single-factor in this context is used in contrast to the term multifactor. In this latter type of experiment the primary objective is to compare the effect of combinations of treatments acting simultaneously on each of the elements. There are some instances in which the distinction between single-factor and multifactor experiments is difficult to make.

In this chapter only designs involving independent observations will be discussed; corresponding designs having correlated observations are considered in the next chapter. The designs in this chapter form a special case of what are called *completely randomized* designs. These form the building

blocks for many other designs; they also provide a standard against which the efficiency of other types of designs is measured.

3.2 Definitions and Numerical Example

A numerical example will illustrate the definitions of terms used in the analysis of a single-factor experiment. The actual analysis of this example will be given in detail. The rationale justifying this analysis will be discussed in the next section.

Table 3.2-1

	Method 1	Method 2	Method 3	
	3	4	6	
	5	4	7	$n = 8$
	2	3	8	
(i)	4	8	6	$k = 3$
	8	7	7	
	4	4	9	
	3	2	10	
	9	5	9	
(ii)†	$T_1 = 38$ $\Sigma X_1^2 = 224$	$T_2 = 37$ $\Sigma X_2^2 = 199$	$T_3 = 62$ $\Sigma X_3^2 = 496$	$G = \Sigma T_j = 137$ $\Sigma(\Sigma X_j^2) = 919$
(iii)	$SS_1 = \Sigma X_1^2 - \frac{T_1^2}{n}$ $= 224 - \frac{38^2}{8}$ $= 43.50$	$SS_2 = \Sigma X_2^2 - \frac{T_2^2}{n}$ $= 199 - \frac{37^2}{8}$ $= 27.88$	$SS_3 = \Sigma X_3^2 - \frac{T_3^2}{n}$ $= 496 - \frac{62^2}{8}$ $= 15.50$	$SS_w = \Sigma SS_j$ $SS_w = 86.88$
(iv)	$\bar{T}_1 = T_1/n$ $= 38/8 = 4.75$	$\bar{T}_2 = T_2/n$ $= 37/8 = 4.62$	$\bar{T}_3 = T_3/n$ $= 62/8 = 7.75$	$\bar{G} = G/nk$ $= 137/24 = 5.71$

† The symbol T_1 is an abbreviation for the more complete notation given by

$$T_1 = \sum_i X_{i1}.$$

Similarly,

$$\Sigma X_1^2 = \sum_i X_{i1}^2.$$

An experimenter is interested in evaluating the effectiveness of three methods of teaching a given course. A group of 24 subjects is available to the experimenter. This group is considered by the experimenter to be the equivalent of a random sample from the population of interest. Three subgroups of 8 subjects each are formed at random; the subgroups are then taught by one of the three methods. Upon completion of the course, each of the subgroups is given a common test covering the material in the course. The resulting test scores are given in part i of Table 3.2-1. The symbol n designates the number of subjects in a subgroup and k the number of methods.

In part ii of this table, the symbol T_j designates the sum of the test scores for the subjects who were taught by method j . For example, T_1 designates the sum of the test scores for the subjects taught by method 1. The symbol G designates the grand total of all observations in the experiment. G is most readily obtained by summing the T_j 's; that is, $G = 38 + 37 + 62 = 137$. The symbol ΣX_j^2 designates the sum of the squares of the observations on the subjects taught by method j . For example, $\Sigma X_1^2 = 3^2 + 5^2 + \dots + 9^2 = 224$.

In part iv, \bar{T}_j designates the mean of the test scores for subjects taught by method j . For example, $\bar{T}_1 = 4.75$ is the mean test score for the subjects under method 1. The symbol \bar{T} designates the grand mean of all test scores. When there is an equal number of subjects in each of the subgroups, $\bar{G} = (\Sigma \bar{T}_j)/k$, where k is the number of subgroups; this relationship does not hold if the number of subjects within each subgroup varies.

The symbol SS_j in part iii designates the sum of squares, or *variation*, of the test scores within the group under method j . By definition, the variation of the observations within method j is

$$(1) \quad SS_j = \sum_i (X_{ij} - \bar{T}_j)^2,$$

i.e., the sum of the squared deviations of the test scores under method j about the mean of subgroup j . This definition is algebraically equivalent to

$$(2) \quad SS_j = \Sigma X_j^2 - \frac{T_j^2}{n},$$

which is a more convenient computational formula. Use of this formula is illustrated in part iii. The symbol SS_w designates the pooled within-method variation (or sum of squares). By definition, SS_w is the sum of the variation within each of the methods,

$$(3) \quad SS_w = \Sigma SS_j.$$

A computationally more convenient formula for SS_w is

$$(4) \quad SS_w = \Sigma (\Sigma X_j^2) - \frac{\Sigma T_j^2}{n}.$$

For the data in the table, (4) is

$$\begin{aligned} SS_w &= 919 - \frac{(38)^2 + (37)^2 + (62)^2}{8} \\ &= 919 - \frac{6657}{8} = 919 - 832.12 = 86.88. \end{aligned}$$

This value for SS_w is the same as that obtained in the table using (3).

The variation (or sum of squares) due to the methods of training is by definition

$$(5) \quad SS_{\text{methods}} = n \Sigma (\bar{T}_j - \bar{G})^2.$$

This statistic measures the extent to which the means for the subgroups differ from the grand mean. It is also a measure of the extent to which the subgroup means differ from one another. In terms of this latter interpretation, (5) may be shown to be algebraically equivalent to

$$(6) \quad SS_{\text{methods}} = \frac{n \sum (\bar{T}_j - \bar{T}_j)^2}{k(k-1)},$$

where the symbol $\bar{T}_j - \bar{T}_j$ designates the difference between a pair of means. For the data in the table, (6) is

$$\begin{aligned} SS_{\text{methods}} &= \frac{8[(4.75 - 4.62)^2 + (4.75 - 7.75)^2 + (4.62 - 7.75)^2]}{3(2)} \\ &= 50.17. \end{aligned}$$

Using (5),

$$\begin{aligned} SS_{\text{methods}} &= 8[(4.75 - 5.71)^2 + (4.62 - 5.71)^2 + (7.75 - 5.71)^2] \\ &= 50.17. \end{aligned}$$

Neither (5) nor (6) is a convenient computational formula for SS_{methods} ; the latter is given by

$$(7) \quad SS_{\text{methods}} = \frac{\sum T_j^2}{n} - \frac{G^2}{nk}.$$

For the data in the table, (7) is numerically

$$\begin{aligned} SS_{\text{methods}} &= \frac{(38)^2 + (37)^2 + (62)^2}{8} - \frac{(137)^2}{24} \\ &= 832.12 - 782.04 = 50.08. \end{aligned}$$

The numerical value for SS_{methods} computed from (5) and (6) involves more rounding errors than does the computational formula (7).

The variation due to experimental error is, by definition, the pooled within-method variation,

$$(8) \quad SS_{\text{error}} = SS_w = \sum SS_j.$$

This statistic measures the sum of the variation within each of the subgroups. Its computational formula is given by (4). The total variation, or total sum of squares, is

$$(9) \quad SS_{\text{total}} = \sum \sum (X_{ij} - \bar{G})^2,$$

the sum of the squared deviation of each observation in the experiment about the grand mean. Its computational formula is

$$(10) \quad SS_{\text{total}} = \sum (\sum X_j^2) - \frac{G^2}{nk}.$$

For the data in the table,

$$\begin{aligned} SS_{\text{total}} &= 919 - \frac{(137)^2}{24} \\ &= 136.96. \end{aligned}$$

From these definitions of SS_{total} , SS_{methods} , and SS_{error} , it may be shown algebraically that

$$(11) \quad \begin{aligned} SS_{\text{total}} &= SS_{\text{methods}} + SS_{\text{error}}, \\ \Sigma \Sigma (X_{ij} - \bar{G})^2 &= n \Sigma (\bar{T}_j - \bar{G})^2 + \Sigma \Sigma (X_{ij} - \bar{T}_j)^2. \end{aligned}$$

The relation (11) describes a *partition*, or division, of the total variation into two additive parts. One part is a function of differences between the mean scores made by the subgroups having different methods of training; the other part is the sum of the variation of scores within subgroups. Numerically,

$$136.96 = 50.08 + 86.88.$$

The partition represented by (11) is basic to the analysis of variance. Its derivation is not difficult. Let

$$a_{ij} = X_{ij} - \bar{T}_j \quad \text{and} \quad b_j = \bar{T}_j - \bar{G}.$$

Then

$$a_{ij} + b_j = X_{ij} - \bar{G},$$

$$\text{and} \quad \Sigma \Sigma (X_{ij} - \bar{G})^2 = \Sigma \Sigma (a_{ij} + b_j)^2 = \Sigma \Sigma a_{ij}^2 + \Sigma \Sigma b_j^2 + 2 \Sigma \Sigma a_{ij} b_j.$$

The term at the extreme right is

$$\Sigma \Sigma a_{ij} b_j = \Sigma b_j (\Sigma_i a_{ij}) = 0,$$

since $\Sigma_i a_{ij} = 0$ for each j . (That is, $\Sigma_i a_{ij}$ is the sum of deviations about the mean of observations in class j .) Since b_j^2 is a constant for all i 's in the same class,

$$\Sigma \Sigma b_j^2 = n \Sigma b_j^2.$$

Hence

$$\begin{aligned} \Sigma \Sigma (X_{ij} - \bar{G})^2 &= \Sigma \Sigma a_{ij}^2 + n \Sigma b_j^2 \\ &= \Sigma \Sigma (X_{ij} - \bar{T}_j)^2 + n \Sigma (\bar{T}_j - \bar{G})^2. \end{aligned}$$

A variance, in the terminology of analysis of variance, is more frequently called a mean square (abbreviated MS). By definition

$$(12) \quad \text{Mean square} = \frac{\text{variation}}{\text{degrees of freedom}} = \frac{SS}{df}.$$

In words, a mean square is the average variation per degree of freedom; this is also the basic definition for a variance. The term mean square is a more general term for the average of squared measures. Hence a variance is actually a special case of a mean square.

The term *degrees of freedom* originates from the geometric representation of problems associated with the determination of sampling distributions for statistics. In this context the term refers to the dimension of the geometric space appropriate in the solution of the problem. The following definition

permits the computation of the degrees of freedom for any source of variation:

$$(13) \text{ Degrees of freedom} = \left(\begin{array}{c} \text{no. of independent} \\ \text{observations on} \\ \text{source of variation} \end{array} \right) - \left(\begin{array}{c} \text{no. of independent} \\ \text{parameters estimated} \\ \text{in computing variation} \end{array} \right).$$

In this context, a statistic may be used either as an estimate of a parameter or as a basic observation in estimating a source of variation. The source of variation being estimated will indicate what role a particular statistic will have in a specified context. More accurately,

$$(13a) \quad df = (\text{no. independent observations}) - (\text{no. linear restraints}).$$

Substituting an \bar{X}_j for a μ_j in the computation of a mean square is equivalent to imposing a linear restraint upon the estimation procedure. The substitution restricts the sum of a set of observations to be a specified number. For example, if the mean of four scores is required to be 10, and if the first three observations are

3, -5, and 20,

then the fourth score must be 22; that is, the total must be 40. Under this restraint on the mean, only three of the four scores are free to vary. Hence the term *freedom*.

In the computation of SS_{methods} , the \bar{T}_j 's are considered to be the basic observations; SS_{methods} is a measure of the variation of the \bar{T}_j 's. Thus, there are k independent observations in SS_{methods} . In the computation of this source of variation, \bar{G} is used as a parameter estimating the mean of the \bar{T}_j 's. Hence one estimate of a parameter is used in the computation of SS_{methods} . Therefore, by (13), the degrees of freedom for this source of variation are $k - 1$.

An alternative, computational definition of the degrees of freedom for a source of variation is

$$(13b) \quad \text{Degrees of freedom} = \left(\begin{array}{c} \text{no. squared} \\ \text{deviations} \end{array} \right) - \left(\begin{array}{c} \text{no. independent points} \\ \text{about which deviations} \\ \text{are taken} \end{array} \right).$$

For example, $SS_{\text{methods}} = n \sum (\bar{T}_j - \bar{G})^2$ involves k squared deviations all taken about the single point \bar{G} . Hence the degrees of freedom are $k - 1$. As another example, $SS_{\text{error}} = \sum SS_j$ involves n squared deviations for each of the k subgroups, or a total of kn squared deviations. Within each subgroup the deviations are taken about the mean \bar{T}_j of that subgroup. Since there are k subgroups, there are k different points about which the deviations are taken. Hence the degrees of freedom are $nk - k$ for SS_{error} .

SS_{error} is the pooled variation within each of the subgroups. The variation within the subgroup j is a measure of the extent to which each of the n observations deviates from the mean of the subgroup, \bar{T}_j . For this source of variation \bar{T}_j is used as an estimate of the mean for population j . Hence

the number of degrees of freedom for the variation within subgroup j is $n - 1$. The degrees of freedom for the pooled within-subgroup variation is the sum of the degrees of freedom for each of the subgroups. If the variation within each of k subgroups has $n - 1$ degrees of freedom, then the total degrees of freedom for k subgroups is $k(n - 1)$, that is, the sum $(n - 1) + (n - 1) + \dots + (n - 1)$ for k terms.

SS_{total} is the variation of the nk independent observations about the grand mean \bar{G} . Here \bar{G} is an estimate of the over-all population mean. Hence SS_{total} is based upon $nk - 1$ degrees of freedom. Corresponding to the partition of the total variation in (11), there is a partition of the total degrees of freedom

$$(14) \quad \begin{aligned} df_{\text{total}} &= df_{\text{methods}} + df_{\text{error}}, \\ kn - 1 &= (k - 1) + (kn - k). \end{aligned}$$

Table 3.2-2 Summary of Analysis of Variance

Source of variation	Sum of squares	Degrees of freedom	Mean square
Between methods	$SS_{\text{methods}} = 50.08$	$k - 1 = 2$	$MS_{\text{methods}} = 25.04$
Experimental error	$SS_{\text{error}} = 86.88$	$kn - k = 21$	$MS_{\text{error}} = 4.14$
Total	$SS_{\text{total}} = 136.96$	$kn - 1 = 23$	

A summary of the statistics used in the analysis of a single-factor experiment is given in Table 3.2-2. The numerical entries are those computed from the data in Table 3.2-1.

Assuming that there is no difference in the effectiveness of the methods of training, as measured by the mean scores on the test, and making additional assumptions which will become explicit in the next section, the statistic

$$(15) \quad F = \frac{MS_{\text{methods}}}{MS_{\text{error}}}$$

has a sampling distribution which is approximated by an F distribution having $k - 1$ degrees of freedom for the numerator and $kn - k$ degrees of freedom for the denominator. Thus the F statistic may be used to test hypotheses about the equality of the population means for the methods. To test the hypothesis that the population means for the test scores are equal, that is, $\mu_1 = \mu_2 = \mu_3$, against a two-tailed alternative hypothesis, the decision rules are as follows:

Reject H_1 when $F_{\text{obs}} > F_{1-\alpha}(k - 1, kn - k)$.

Otherwise do not reject H_1 .

For the data in Table 3.2-2, $F_{\text{obs}} = 25.04/4.14 = 6.05$. Critical values for $\alpha = .05$ and $\alpha = .01$ are, respectively, $F_{.95}(2, 21) = 3.47$ and $F_{.99}(2, 21) = 5.78$. In this case F_{obs} exceeds the critical value for $\alpha = .01$. Hence the

data do not support the hypothesis that the population means are equal. Inspection of the means in Table 3.2-1 indicates that method 3 has the largest mean.

The experiment represented by the data in Table 3.2-1 is a special case of a single-factor experiment. For this case, the experimental variable is the method of training. In the general case the term treatment will be used

Table 3.2-3 General Notation

	Treatment 1	...	Treatment j	...	Treatment k	
(i)	X_{11}		X_{1j}		X_{1k}	
	X_{21}		X_{2j}		X_{2k}	
	\vdots		\vdots		\vdots	
	\vdots		\vdots		\vdots	
	X_{i1}		X_{ij}		X_{ik}	
	\vdots		\vdots		\vdots	
	X_{n1}		X_{nj}		X_{nk}	
(ii)	T_1	...	T_j	...	T_k	G
	ΣX_1^2		ΣX_j^2		ΣX_k^2	$\Sigma(\Sigma X_j^2)$
	\bar{T}_1		\bar{T}_j		\bar{T}_k	\bar{G}
(iii)	$SS_{\text{treat}} = \frac{\Sigma T_j^2}{n} - \frac{G^2}{kn}$					$df_{\text{treat}} = k - 1$
	$SS_{\text{error}} = \Sigma(\Sigma X_j^2) - \frac{\Sigma T_j^2}{n}$					$df_{\text{error}} = kn - k$
	$SS_{\text{total}} = \Sigma(\Sigma X_j^2) - \frac{G^2}{kn}$					$df_{\text{total}} = kn - 1$
(iv)	<i>Computational symbols</i>					<i>Sums of squares in terms of computational symbols</i>
	(1) = G^2/kn					$SS_{\text{treat}} = (3) - (1)$
	(2) = $\Sigma(\Sigma X_j^2)$					$SS_{\text{error}} = (2) - (3)$
	(3) = $(\Sigma T_j^2)/n$					$SS_{\text{total}} = (2) - (1)$

interchangeably with the terms experimental variable, experimental condition, or whatever it is that distinguishes the manner in which the subgroups are handled (treated) in the experiment. The elements assigned to a treatment constitute what will be called a treatment class. A general notation for the observed data in a single-factor experiment having n observations in each treatment class is given in part i of Table 3.2-3. For example, an observation on the element i in treatment class j is designated X_{ij} . Notation for the totals required in the computation of the sums of squares appears

in part ii. For example, T_j designates the sum of all observations in treatment class j , and ΣX_j^2 designates the sum of the squares of the observations in treatment class j . Both sums are generally obtained in a single operation on a calculator.

In part iii the computational formulas for the sums of squares used in the analysis of variance and the associated degrees of freedom are given. A convenient method for summarizing the computational formulas is given in part (iv) in terms of what may be called computational symbols. For example, the symbol (2) designates the numerical value of $\Sigma(\Sigma X_j^2)$.

The degrees of freedom for a sum of squares may be computed directly from the computational formula by means of the following rule: Count the number of quantities which are squared in a term. Then replace this term by this number in the computational formula. For example, in the term $(\Sigma T_j^2)/n$ there are k quantities that are squared (T_1, T_2, \dots, T_k). In the

Table 3.2-4 General Form of Summary Data

Source	SS	df	MS	F
Treatments	SS_{treat}	$k - 1$	MS_{treat}	$F = \frac{MS_{\text{treat}}}{MS_{\text{error}}}$
Experimental error	SS_{error}	$kn - k$	MS_{error}	
Total	SS_{total}	$kn - 1$		

term G^2/kn there is just one term which is squared, namely, G . Hence the degrees of freedom of the sum of squares defined by $[(\Sigma T_j^2)/n - G^2/kn]$ are $k - 1$. As another example, in the term $\Sigma(\Sigma X_j^2)$ there are kn terms that are squared, namely, each of the kn individual observations. Hence the sum of squares defined by $[\Sigma(\Sigma X_j^2) - G^2/kn]$ has $kn - 1$ degrees of freedom.

The general form used in summarizing the analysis of variance for a single-factor experiment is given in Table 3.2-4. The F statistic is used in testing the hypothesis that $\mu_1 = \mu_2 = \dots = \mu_k$ against the equivalent of a two-tailed alternative hypothesis. If this hypothesis is rejected, additional tests are required for more detailed information about which means are different from the others. Specialized tests for comparing individual means with each other are discussed in later sections.

The formal method for testing statistical hypotheses requires that the level of significance of a test be set in advance of obtaining the data. Convention in the analysis of variance is somewhat opposed to this procedure. The value of F_{obs} is generally compared with tabled critical values, and the outcome is described in terms of the statement: F_{obs} exceeds a specified percentile point (usually the .95 or the .99 percentile points). The choice of the level of significance is thus in part determined by the observed data. This procedure is not objectionable for purposes of estimating the probability

of the observed outcome in terms of an assumed underlying sampling distribution. However, this procedure does not permit the evaluation of the power associated with the test.

3.3 Structural Model for Single-factor Experiment—Model I

Suppose all elements in a specified basic population are given treatment 1. After the treatment, the elements are measured on a criterion related to the effectiveness of the treatment. Assume that the distribution of the resulting measurements is approximately normal in form, with parameters

Table 3.3-1 Parameters of Populations Corresponding to Treatments in the Experiment

Treatment	Population mean	Population variance	Treatment effect
1	μ_1	σ_1^2	$\tau_1 = \mu_1 - \mu$
2	μ_2	σ_2^2	$\tau_2 = \mu_2 - \mu$
.	.	.	.
.	.	.	.
.	.	.	.
j	μ_j	σ_j^2	$\tau_j = \mu_j - \mu$
.	.	.	.
.	.	.	.
j'	$\mu_{j'}$	$\sigma_{j'}^2$	$\tau_{j'} = \mu_{j'} - \mu$
.	.	.	.
.	.	.	.
$k = K$	μ_k	σ_k^2	$\tau_k = \mu_k - \mu$
	Grand mean = μ		$\bar{\tau} = 0$

μ_1 and σ_1^2 . Thus, μ_1 and σ_1^2 are the parameters of a population of measurements that would exist if treatment 1 were administered. This potential population will be designated as that corresponding to treatment 1. Suppose, instead, that treatment 2 is administered to all the elements and then measurements on the same criterion of effectiveness are made. Assume that the resulting distribution is approximately normal in form, with parameters μ_2 and σ_2^2 .

Corresponding to each treatment about which the experimenter seeks to make inferences, there is assumed to be a population of approximately normally distributed criterion measures. The number of such populations is equal to the number of treatments. Assume that the number of treatments is K . In the experiment, data are obtained on k of the possible K treatments. If $k = K$, then observed data are available on all treatments in the domain to which inferences are to be made. If k is less than K , then observed data are

available on only some of the treatments about which inferences are to be drawn. In this section the case in which $k = K$ will be considered. Other cases are considered in the next section.

The parameters defining the treatment populations are summarized in Table 3.3-1. When $k = K$, the grand mean μ is

$$(1) \quad \mu = \frac{\sum \mu_j}{k}.$$

The effect of treatment j , designated τ_j , is the difference between the mean for treatment j and the grand mean of the population means,

$$(2) \quad \tau_j = \mu_j - \mu.$$

Thus τ_j is a parameter which measures the degree to which the mean for treatment j differs from the mean of all other relevant population means. Since the sum of the deviations about the mean of a set is zero, $\sum \tau_j = 0$; hence the mean of the τ_j 's, designated $\bar{\tau}$, is equal to zero.

Let X_{ij} be the criterion measure on a randomly selected element i in treatment population j . The following structural model is assumed for this measurement,

$$(3) \quad X_{ij} = \mu + \tau_j + \varepsilon_{ij},$$

where μ = grand mean of treatment populations,

τ_j = effect of treatment j ,

ε_{ij} = experimental error.

The term μ is constant for all measurements in all treatment populations. The effect τ_j is constant for all measurements within population j ; however, a different value, say, $\tau_{j'}$, is associated with population j' , where j' represents some treatment other than j . The experimental error ε_{ij} represents all the uncontrolled sources of variance affecting individual measurements; this effect is unique for each of the elements i in the basic population. This effect is further assumed to be independent of τ_j .

Since both μ and τ_j are constant for all measurements within population j , the only source of variance for these measurements is that due to experimental error. Thus

$$(4) \quad \sigma_j^2 = \sigma_{\varepsilon_j}^2,$$

where $\sigma_{\varepsilon_j}^2$ designates the variance due to experimental error for the measurements within treatment population j . If $X_{ij'}$ represents a measurement in population j' , then (3) takes the form

$$(5) \quad X_{ij'} = \mu + \tau_{j'} + \varepsilon_{ij'}.$$

The variance within population j' is due solely to the experimental error; hence

$$(6) \quad \sigma_{j'}^2 = \sigma_{\varepsilon_{j'}}^2,$$

If the elements (subjects, animals) that are observed under each of the treatment conditions are assigned at random to these conditions, one has some degree of assurance that the error effects will be independent of the treatment effects. Hence the importance of randomization in design problems. The elements, in this context, are called the experimental units. Random assignment of the experimental units to the experimental conditions tends to make the unique effects of the units per se independent of the treatment effects.

The experimental error measures all uncontrolled effects which are not related to the treatments. As such, the experimental error is the combined effect of many random variables that are independent of the treatment effect. Under these conditions, it is reasonable to assume that the distribution of the ε_{ij} 's within population j will be approximately normal in form, with expected value $\mu_{\varepsilon_j} = 0$ and variance $\sigma_{\varepsilon_j}^2$. If the sources of experimental error are comparable in each of the treatment populations, it is also reasonable to assume that

$$(7) \quad \sigma_{\varepsilon_j}^2 = \sigma_{\varepsilon_{j'}}^2.$$

This last relationship may be written in more general form as

$$(8) \quad \sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2 = \cdots = \sigma_{\varepsilon_k}^2 = \sigma_{\varepsilon}^2,$$

where σ_{ε}^2 is the variance due to experimental error within any of the treatment populations.

To summarize the assumptions underlying the structural model (3), a measurement X_{ij} is expressed as the sum of three components: (1) a component μ which is constant for all treatments and all elements; (2) a component τ_j which is constant for all elements within a treatment population but may differ for different treatment populations; (3) a component ε_{ij} , independent of τ_j , and distributed as $N(0, \sigma_{\varepsilon}^2)$ within each treatment population. This structural model is called the *fixed-constants* model, or model I. The component μ in model I is actually an unknown constant; the component τ_j is a systematic or fixed component which depends upon the difference between the means of the treatment populations; the component ε_{ij} is a random component (or a random variable) depending upon uncontrolled sources of variances assumed to be drawn randomly from a population in which the distribution is $N(0, \sigma_{\varepsilon}^2)$.

A parameter indicating the extent to which the treatment effects differ is

$$(9) \quad \sigma_{\tau}^2 = \frac{\Sigma \tau_j^2}{k-1} = \frac{\Sigma (\mu_j - \mu)^2}{k-1}.$$

An equivalent definition in terms of differences between treatment effects is

$$(10) \quad \sigma_{\tau}^2 = \frac{\Sigma (\tau_j - \tau_{j'})^2}{k(k-1)} = \frac{\Sigma (\mu_j - \mu_{j'})^2}{k(k-1)},$$

where the summation is over all the different possible pairs of means; there are $k(k-1)/2$ such distinct pairs. When the treatment effects are equal, i.e., when $\tau_1 = \tau_2 = \dots = \tau_k$, σ_τ^2 will be zero. The larger the differences between the τ 's the larger will be σ_τ^2 . Thus the hypothesis specifying that $\sigma_\tau^2 = 0$ is equivalent to the hypothesis that specifies $\tau_1 = \tau_2 = \dots = \tau_k$ or $\mu_1 = \mu_2 = \dots = \mu_k$.

Statistics useful in estimating the parameters in Table 3.3-1 are summarized in Table 3.3-2. Since model I assumes that the population

Table 3.3-2 Estimates of Parameters of Treatment Populations

Sample size	Treatment	Sample mean	Sample variance	Treatment effect
n	1	\bar{T}_1	s_1^2	$t_1 = \bar{T}_1 - \bar{G}$
n	2	\bar{T}_2	s_2^2	$t_2 = \bar{T}_2 - \bar{G}$
.
.
.
n	j	\bar{T}_j	s_j^2	$t_j = \bar{T}_j - \bar{G}$
.
.
n	j'	$\bar{T}_{j'}$	$s_{j'}^2$	$t_{j'} = \bar{T}_{j'} - \bar{G}$
.
.
n	k	\bar{T}_k	s_k^2	$t_k = \bar{T}_k - \bar{G}$
		Grand mean = \bar{G}	$\bar{i} = 0$	

variances are all equal to σ_e^2 , the best estimate of this parameter is the pooled within-class sample variance

$$(11) \quad s_{\text{pooled}}^2 = \frac{\sum s_j^2}{k} = \text{MS}_{\text{error}}.$$

For this design the pooled within-class variance is designated MS_{error} . The latter is an unbiased estimate of σ_e^2 ; that is,

$$E(\text{MS}_{\text{error}}) = \sigma_e^2.$$

In terms of the structural model in (3), the treatment means for the samples of n observations may be expressed as

$$(12) \quad \begin{aligned} \bar{T}_1 &= \mu + \tau_1 + \bar{\epsilon}_1, \\ \bar{T}_2 &= \mu + \tau_2 + \bar{\epsilon}_2, \\ &\dots \dots \dots, \\ \bar{T}_k &= \mu + \tau_k + \bar{\epsilon}_k, \end{aligned}$$

where $\bar{\epsilon}_j$ is the mean experimental error for a sample of n observations within treatment class j . For random samples, it is reasonable to assume that $E(\bar{\epsilon}_j) = \text{constant}$. Without any loss in generality, this constant may be assumed to be zero. Therefore,

$$(13) \quad \begin{aligned} E(\bar{T}_1) &= \mu + \tau_1, \\ E(\bar{T}_2) &= \mu + \tau_2, \\ &\dots \dots \dots, \\ E(\bar{T}_k) &= \mu + \tau_k. \end{aligned}$$

The first relation in (13) may be interpreted as follows: If an infinite number of random samples of size n are given treatment 1 and the statistic \bar{T}_1 is computed for each of the samples, the distribution of the resulting statistic would have an expected value (or mean) equal to $\mu + \tau_1$.

The statistic t_j in Table 3.3-2 may be represented as

$$(14) \quad \begin{aligned} t_j &= \bar{T}_j - \bar{G} = (\mu + \tau_j + \bar{\epsilon}_j) - (\mu + \bar{\epsilon}) \\ &= \tau_j + \bar{\epsilon}_j - \bar{\epsilon}. \end{aligned}$$

where $\bar{\epsilon}$ is the mean of the $\bar{\epsilon}_j$'s. The expected value of t_j is equal to τ_j , since the expected value of the quantity $\bar{\epsilon}_j - \bar{\epsilon}$ equals 0.

A measure of the degree to which the sample means for the various treatments differ is provided by the statistic

$$(15) \quad MS_{\text{treat}} = \frac{n \sum (\bar{T}_j - \bar{G})^2}{k - 1} = \frac{n \sum t_j^2}{k - 1}.$$

Inspection of the right-hand side of the relation (14) indicates that differences among the \bar{T} 's depend upon differences among the τ_j 's and differences among the $\bar{\epsilon}_j$'s. Since the τ_j 's and the $\bar{\epsilon}_j$'s are assumed to be independent, the variance of the \bar{T} 's is the sum of the variance of the τ 's and the variance of the $\bar{\epsilon}$'s. Therefore n times the variance of the \bar{T} 's has the expected value

$$(16) \quad \begin{aligned} E(MS_{\text{treat}}) &= \frac{n \sum \tau_j^2}{k - 1} + \frac{n \sum \bar{\epsilon}_j^2}{k - 1} \\ &= n \sigma_\tau^2 + n \sigma_{\bar{\epsilon}}^2. \end{aligned}$$

Since the variance of the mean of n observations is n times the variance of the individual observations, using the relationship that $\sigma_{\bar{\epsilon}}^2 = n \sigma_\epsilon^2$ permits (16) to be written as

$$(17) \quad E(MS_{\text{treat}}) = n \sigma_\tau^2 + \sigma_\epsilon^2.$$

Thus, when $\tau_1 = \tau_2 = \dots = \tau_k$, the term σ_τ^2 is equal to zero and the expected value of MS_{treat} is σ_ϵ^2 . This implies that MS_{treat} is an unbiased estimate of the variance due to experimental error when there are no differences among the treatment effects.

A more general approach to the result obtained in (17) is to start from the relation

$$(18) \quad \bar{T}_j - \bar{T}_{j'} = (\tau_j - \tau_{j'}) + (\bar{\varepsilon}_j - \bar{\varepsilon}_{j'}).$$

This relation is obtained from (12). In this form, a difference between two treatment means estimates a difference between two treatment effects and a difference between two average error effects. Further, $\bar{T}_j - \bar{T}_{j'}$ provides an unbiased estimate of $\tau_j - \tau_{j'}$. MS_{treat} is readily expressed in terms of all possible pairs of differences among the \bar{T} 's. Similarly σ_τ^2 and σ_ε^2 may be expressed in terms of the differences on the right-hand side of (18). In the derivation of the expected values for the mean square due to treatments, the more direct approach is by means of relations analogous to (18).

Table 3.3-3 Expected Values of Mean Squares
(n Observations per Treatment)

Source	df	MS	E(MS)
Treatments	$k - 1$	MS_{treat}	$\sigma_\varepsilon^2 + n\sigma_\tau^2$
Experimental error	$kn - k$	MS_{error}	σ_ε^2

The expected values of the mean squares which are used in the analysis of variance are summarized in Table 3.3-3. In terms of these expected values, since the sampling distributions of MS_{treat} and MS_{error} are independent,

$$E\left(\frac{MS_{\text{treat}}}{MS_{\text{res}}}\right) = \frac{\sigma_\varepsilon^2 + n\sigma_\tau^2}{\sigma_\varepsilon^2}.$$

When $\sigma_\tau^2 = 0$, the expected values of the numerator and denominator of this ratio are both equal to σ_ε^2 . Hence numerator and denominator are independent, unbiased estimates of σ_ε^2 when $\sigma_\tau^2 = 0$. The independence of the estimates follows from the fact that the sampling distribution of the within-class variance is independent of the class means. MS_{error} is obtained from the within-class data, whereas MS_{treat} is obtained from the class means.

Thus, when $\sigma_\tau^2 = 0$, the sampling distribution of the statistic

$$(19) \quad F = \frac{MS_{\text{treat}}}{MS_{\text{error}}}$$

is an F distribution having $k - 1$ degrees of freedom for the numerator and $kn - k$ degrees of freedom for the denominator. This F statistic may be used to test the hypothesis that $\sigma_\tau^2 = 0$, which is equivalent to the hypothesis that $\tau_1 = \tau_2 = \cdots = \tau_k$. The assumptions under which this test is valid are those underlying model I; these assumptions were discussed in connection with the structural model. The magnitude of type 1 error is not seriously affected if the distributions depart moderately from normality or if the population variances depart moderately from equality; i.e., the test is robust

with respect to the assumptions of normality of distribution and homogeneity of error variance. A study of the effect of lack of homogeneity of error variance is given in the work of Box (1954).

When $\sigma_\tau^2 \neq 0$, the expected value of the F statistic will be greater than 1.00 by an amount which depends upon the magnitude of σ_τ^2 . Thus, if the F ratio is larger than 1.00 by an amount having low probability when $\sigma_\tau^2 = 0$, the inference is that $\sigma_\tau^2 \neq 0$.

The notation $\text{ave}(\text{MS})$, read average value of a mean square, is used by some authors in place of the notation $E(\text{MS})$. Also the notation σ_τ^2 is sometimes reserved for use in connection with model II, and no special symbol is used to designate the quantity $n(\sum \tau_j^2)/(k-1)$. Thus, for model I

$$E(\text{MS}_{\text{treat}}) = \sigma_e^2 + \frac{n \sum \tau_j^2}{k-1}.$$

As long as the assumptions underlying the model are made explicit, no ambiguity will result in the use of the symbol σ_τ^2 in model I.

3.4 Structural Model for Single-factor Experiment—Model II (Variance Component Model)

One of the basic assumptions underlying model I is that all the treatments about which inferences are to be made are included in the experiment. Thus, if the experiment were to be replicated, the same set of treatments would be included in each of the replications. Model I is usually the most appropriate for a single-factor experiment. If, however, the k treatments that are included in a given experiment constitute a random sample from a collection of K treatments, where k is small relative to K , then upon replication a different random sample of k treatments will be included in the experiment.

Model II covers this latter case. The structural equation for model II has the same form as that of model I, namely,

$$(1) \quad X_{ij} = \mu + \tau_j + \varepsilon_{ij}.$$

However, the assumptions underlying this model are different. The term μ is still assumed to be constant for all observations; the term ε_{ij} is still assumed to have the distribution $N(0, \sigma_e^2)$ for all treatments; but the term τ_j is now considered to be a random variable. Whether or not a given treatment is included in an experiment depends upon a random sampling procedure. Within the sample of n observations in treatment class j , τ_j is considered to be a constant. The distribution of τ_j is assumed to be $N(0, \sigma_\tau^2)$.

The estimates of the parameters in (1) have the same form and are numerically equal to estimates made from data which assume model I. From the point of view of the computation of estimates of the various effects, the two

models do not lead to different procedures. For samples of size n , the expected values of the mean squares of the estimators are as follows:

Source	df	MS	E(MS)
Treatments	$k - 1$	MS_{treat}	$\sigma_e^2 + n \left(1 - \frac{k}{K}\right) \sigma_\tau^2$
Experimental error	$kn - k$	MS_{error}	σ_e^2

If k is small relative to K , then the coefficient $1 - (k/K)$ becomes very close to 1.00 and the expected value of the mean square for treatments becomes $\sigma_e^2 + n\sigma_\tau^2$. This latter value was the one that was obtained under model I. The test on the hypothesis $\sigma_\tau^2 = 0$ is identical with that made under model I. In more complex designs, the F ratios for analogous tests do not have identical forms.

3.5 Methods for Deriving Estimates and Their Expected Values

Under model I, the structural equation for an observation has the form

$$(1) \quad X_{ij} = \mu + \tau_j + \varepsilon_{ij},$$

where μ is a constant for all observations, τ_j is a constant (or fixed variable) for observations under treatment j , and ε_{ij} is a random variable independent of τ_j having the distribution $N(0, \sigma_e^2)$ for all treatments. Thus, there are k fixed variables and one random variable in the model. This model is similar to that underlying a k -variate multiple regression equation; hence the methods for the solution of estimation problems in multiple regression are applicable. The method of least squares, which is used in the solution of multiple regression problems, will yield unbiased minimum-variance estimators when applied to estimation problems posed by model I. (The usual multiple regression problem involves correlated predictors. In this case, the predictors are uncorrelated. The latter restriction simplifies the computational procedures, but the underlying principles are identical.)

The generalization of model I forms a major part of what is known as classical analysis of variance; estimation problems in this area may be handled by means of least-squares analysis or by means of maximum-likelihood methods. The two approaches lead to identical results for this model.

To illustrate the least-squares approach to (1), let m , t_1 , t_2 , \dots , t_k , and e_{ij} designate the respective least-squares estimators of μ , τ_1 , τ_2 , \dots , τ_k , and ε_{ij} . The method of least squares minimizes the quantity

$$(2) \quad \sum_i \sum_j e_{ij}^2 = \sum_i \sum_j (X_{ij} - m - t_j)^2 = \text{minimum.}$$

By taking partial derivatives with respect to the estimators m , t_1 , t_2 , \dots , t_k (this is the procedure used to obtain the normal equations in the multiple

regression problem) and setting these derivatives equal to zero, one obtains the equations

$$\begin{aligned}
 (3) \quad & knm + n\sum t_j - G = 0, \\
 & nm + nt_1 - T_1 = 0, \\
 & nm + nt_2 - T_2 = 0, \\
 & \dots\dots\dots, \\
 & nm + nt_k - T_k = 0.
 \end{aligned}$$

There are $k + 1$ equations and $k + 1$ unknowns; however, the $k + 1$ equations are not independent, since the first equation may be obtained by summing the last k equations. A unique solution to these equations requires a restriction on the variables. Since $\sum \tau_j = 0$, a reasonable restriction is that $\sum t_j = 0$. The solutions to (3) now take the form

$$\begin{aligned}
 (4) \quad & m = \frac{G}{kn} = \bar{G}, \\
 & t_1 = \frac{T_1 - n\bar{G}}{n} = \bar{T}_1 - \bar{G}, \\
 & t_2 = \frac{T_2 - n\bar{G}}{n} = \bar{T}_2 - \bar{G}, \\
 & \dots\dots\dots, \\
 & t_k = \frac{T_k - n\bar{G}}{n} = \bar{T}_k - \bar{G}.
 \end{aligned}$$

Each of these estimators is an unbiased estimate of the corresponding parameter in (1). Thus,

$$E(m) = E(\bar{G}) = \mu; E(t_1) = E(\bar{T}_1 - \bar{G}) = \tau_1; E(t_2) = E(\bar{T}_2 - \bar{G}) = \tau_2.$$

Although the expected value of t_1 is the parameter τ_1 , in terms of the parameters of the model, t_j has the form

$$\begin{aligned}
 (5) \quad t_j &= \bar{T}_j - \bar{G} = \left(\mu + \tau_j + \frac{\sum_i \varepsilon_{ij}}{n} \right) - \left(\mu + \frac{n\sum \tau_j}{kn} + \frac{\sum \sum \varepsilon_{ij}}{kn} \right) \\
 &= \tau_j + \bar{\varepsilon}_j - \bar{\varepsilon},
 \end{aligned}$$

where $\bar{\varepsilon}_j = (\sum_i \varepsilon_{ij})/n$ and $\bar{\varepsilon} = (\sum \sum \varepsilon_{ij})/kn$. Thus,

$$(6) \quad MS_{\text{treat}} = \frac{n\sum t_j^2}{k-1} = \frac{n\sum (\tau_j + \bar{\varepsilon}_j - \bar{\varepsilon})^2}{k-1}.$$

Since the τ 's and the ε 's are independent, the expected value of MS_{treat} will be n times the sum of the mean squares of the parameters on the right-hand side of (6), that is, $n(\sigma_\tau^2 + \sigma_\varepsilon^2) = n\sigma_\tau^2 + n\sigma_\varepsilon^2$. This is a general property of least-squares estimators: the expected value of the mean square of an estimator is the sum of the respective mean squares of the parameters estimated.

The term

$$\begin{aligned}
 (7) \quad e_{ij} &= X_{ij} - m - t_j = X_{ij} - \bar{G} - (\bar{T}_j - \bar{G}) \\
 &= X_{ij} - \bar{T}_j = (\mu + \tau_j + \varepsilon_{ij}) - (\mu + \tau_j + \bar{\varepsilon}_j) \\
 &= \varepsilon_{ij} - \bar{\varepsilon}_j.
 \end{aligned}$$

Thus

$$(8) \quad E(e_{ij}^2) = E\left(\frac{\sum \sum (X_{ij} - \bar{T}_j)^2}{kn - k}\right) = \sigma_e^2.$$

Model II, which has more than one random variable, does not permit the direct application of least-squares methods. Instead, maximum-likelihood methods are used to obtain estimators and the expected values of the mean squares of these estimators. General methods for obtaining the expected values for mean squares under model II are discussed by Crump (1951).

In experimental designs which are discussed in later chapters, a mixture of model I and model II often occurs. The expected values of the mean squares in designs having a mixed model depend upon the randomization restrictions which are imposed upon the variables. When different sets of restrictions are imposed, different expected values for the mean squares are obtained.

3.6 Comparisons among Treatment Means

A *comparison* or *contrast* between two treatment means is by definition the difference between the two means, disregarding the algebraic sign. Thus, $\bar{T}_1 - \bar{T}_2$ defines a comparison between these two means. Comparisons among three treatment means can be made in several different ways. Each mean may be compared or contrasted with each of the other means, giving rise to the following differences:

$$\bar{T}_1 - \bar{T}_2, \quad \bar{T}_1 - \bar{T}_3, \quad \bar{T}_2 - \bar{T}_3.$$

Another set of comparisons may be obtained by averaging two of the three means and then comparing this average with the third mean. This procedure gives rise to the following differences:

$$\frac{\bar{T}_1 + \bar{T}_2}{2} - \bar{T}_3, \quad \frac{\bar{T}_1 + \bar{T}_3}{2} - \bar{T}_2, \quad \frac{\bar{T}_2 + \bar{T}_3}{2} - \bar{T}_1.$$

In general, a *comparison* or *contrast* among three means is an expression of the form

$$(1) \quad c_1 \bar{T}_1 + c_2 \bar{T}_2 + c_3 \bar{T}_3, \quad \text{where } \sum c_j = 0.$$

[Any expression having the form

$$w_1 \bar{T}_1 + w_2 \bar{T}_2 + w_3 \bar{T}_3,$$

where there is no restriction on the w 's, is called a *linear function* of the \bar{T} 's. A comparison, as defined in (1), is a specialized linear function of the \bar{T} 's.]

For example, if $c_1 = 1$, $c_2 = -1$, and $c_3 = 0$, (1) becomes

$$1\bar{T}_1 + (-1)\bar{T}_2 + (0)\bar{T}_3 = \bar{T}_1 - \bar{T}_2.$$

If $c_1 = 0$, $c_2 = 1$, and $c_3 = -1$, then (1) becomes

$$\bar{T}_2 - \bar{T}_3.$$

If $c_1 = \frac{1}{2}$, $c_2 = -1$, and $c_3 = \frac{1}{2}$, then (1) becomes

$$\left(\frac{1}{2}\right)\bar{T}_1 + (-1)\bar{T}_2 + \left(\frac{1}{2}\right)\bar{T}_3 = \frac{\bar{T}_1 + \bar{T}_3}{2} - \bar{T}_2.$$

The general definition of a comparison or contrast among k means has the form

$$(2) \quad c_1\bar{T}_1 + c_2\bar{T}_2 + \cdots + c_k\bar{T}_k, \quad \text{where } \sum c_j = 0.$$

The expression (2) is called a linear combination or weighted sum of the means. If the sum of the weights in this expression is equal to zero, then the linear combination is a comparison or contrast among the means. When the number of observations in each treatment class is equal to n , it is more convenient to work with totals rather than means. A comparison along the treatment totals is defined by

$$(3) \quad c_1T_1 + c_2T_2 + \cdots + c_kT_k, \quad \text{where } \sum c_j = 0.$$

A comparison between the first and second treatments is obtained by setting $c_1 = 1$, $c_2 = -1$, and setting all other c 's equal to zero.

Orthogonal comparisons will be defined by means of a numerical example. Consider the following comparisons:

$$C_1 = (-3)T_1 + (-1)T_2 + (1)T_3 + (3)T_4,$$

$$C_2 = (1)T_1 + (-1)T_2 + (-1)T_3 + (1)T_4,$$

$$C_3 = (1)T_1 + (1)T_2 + (-1)T_3 + (-1)T_4.$$

For comparisons C_1 and C_2 , the sum of the products of corresponding coefficients is

$$(-3)(1) + (-1)(-1) + (1)(-1) + (3)(1) = 0.$$

For comparisons C_1 and C_3 , the sum of the products of corresponding coefficients is

$$(-3)(1) + (-1)(1) + (1)(-1) + (3)(-1) = -6.$$

Two comparisons are orthogonal if the sum of the products of corresponding coefficients is equal to zero. Thus, the comparisons C_1 and C_2 are orthogonal, but the comparisons C_1 and C_3 are not orthogonal. To find out whether or not C_2 is orthogonal to C_3 , one forms the sum of products of corresponding coefficients,

$$(1)(1) + (-1)(1) + (-1)(-1) + (1)(-1) = 0.$$

Since this sum is equal to zero, the two comparisons are orthogonal. The concept of orthogonality in this context is analogous to the concept of nonoverlapping or uncorrelated sources of variation.

A component of the sum of squares for treatments is defined by the expression

$$\frac{(c_1 T_1 + c_2 T_2 + \cdots + c_k T_k)^2}{n(c_1^2 + c_2^2 + \cdots + c_k^2)}, \quad \text{where } \Sigma c = 0.$$

(It is assumed that there are n observations in each treatment total.) In words, a component of a sum of squares is the square of a comparison divided by n times the sum of the squares of the coefficients in the comparison. A component of a sum of squares has one degree of freedom because it basically represents the squared difference between two observations, each observation being the weighted average of treatment means. A sum of squares based upon two basic observations has one degree of freedom.

Two components of the sum of squares for treatments are orthogonal if the comparisons in these components are orthogonal. Any treatment sum of squares having $k - 1$ degrees of freedom can be divided into $k - 1$ orthogonal components; there are usually many ways in which this can be done. A treatment sum of squares can also be subdivided into a relatively large number of nonorthogonal components. Orthogonal components are additive; i.e., the sum of the parts equals the whole. Each component in an orthogonal set covers a different portion of the total variation. Non-orthogonal components are not additive in this sense.

The computation and interpretation of the components of the sums of squares for treatments will be illustrated by means of the numerical example used in Table 3.2-1. For these data,

$$n = 8, \quad k = 3, \quad T_1 = 38, \quad T_2 = 37, \quad T_3 = 62, \\ SS_{\text{methods}} = 50.08, \quad MS_{\text{error}} = 4.14.$$

Consider the following comparisons:

$$C_1 = (1)T_1 + (-1)T_2 + (0)T_3 = T_1 - T_2 = 38 - 37 = 1,$$

$$C_2 = (1)T_1 + (1)T_2 + (-2)T_3 = T_1 + T_2 - 2T_3 = 38 + 37 - 124 = -49.$$

C_1 represents a comparison between methods 1 and 2, method 3 being disregarded. This comparison is used in testing the hypothesis that $\mu_1 - \mu_2 = 0$. C_2 represents a comparison between method 3 and the average of methods 1 and 2. This comparison is used in testing the hypothesis that $[(\mu_1 + \mu_2)/2] - \mu_3 = 0$. C_1 and C_2 are orthogonal, since the sum of the products of corresponding coefficients is zero; that is,

$$(1)(1) + (-1)(1) + (0)(-2) = 0.$$

The component of the sum of squares corresponding to C_1 is

$$SS_{C_1} = \frac{(T_1 - T_2)^2}{n[(1)^2 + (-1)^2 + (0)^2]} = \frac{(1)^2}{8(2)} = .0625.$$

This component is interpreted as follows: Of the total variation among the methods of training, $SS_{\text{methods}} = 50.08$, that part which is due to the

difference between methods 1 and 2 is $SS_{C_1} = .0625$. The component corresponding to C_2 is

$$SS_{C_2} = \frac{(T_1 + T_2 - 2T_3)^2}{8[(1)^2 + (1)^2 + (-2)^2]} = \frac{(-49)^2}{8(6)} = 50.02.$$

Thus, of the total variation between the methods, that part which is due to the difference between method 3 and the other two methods combined is $SS_{C_2} = 50.02$. Method 3 appears to be clearly different from the other two methods; there is very little difference between methods 1 and 2.

Since a component of a sum of squares has one degree of freedom, there is no difference between a sum of squares or a mean square for a component. The expected value of the sum of squares for C_1 is

$$E(SS_{C_1}) = \sigma_e^2 + \frac{n}{2}(\tau_1 - \tau_2)^2.$$

Under the hypothesis that $\tau_1 - \tau_2 = 0$, $E(SS_{C_1}) = \sigma_e^2$. Hence the ratio

$$\frac{E(SS_{C_1})}{E(MS_{\text{error}})} = \frac{\sigma_e^2}{\sigma_e^2} = 1.00, \quad \text{when } \tau_1 - \tau_2 = 0.$$

Thus the statistic

$$F = \frac{SS_{C_1}}{MS_{\text{error}}}$$

may be used in a test of the hypothesis that $\tau_1 - \tau_2 = 0$. This F ratio has one degree of freedom for the numerator and $kn - k$ degrees of freedom for the denominator. For these data $F_{.95}(1,21) = 4.32$. The numerical value of F_{obs} is

$$F_{\text{obs}} = \frac{.0625}{4.14}.$$

This F ratio is less than 1.00 and hence does not exceed the critical value 4.32. Thus, the observed data do not contradict the hypothesis that $\tau_1 - \tau_2 = 0$.

The hypothesis that $[(\tau_1 + \tau_2)/2] - \tau_3 = 0$ is tested by use of the statistic

$$F = \frac{SS_{C_2}}{MS_{\text{error}}} = \frac{50.02}{4.14} = 12.08.$$

The critical value is again $F_{.95}(1,21) = 4.32$. Thus, the observed data contradict the hypothesis that method 3 is no different from the average effect of methods 1 and 2.

The comparison used in a test of the hypothesis that $\tau_1 - \tau_3 = 0$ is

$$T_1 - T_3 = 38 - 62 = -24.$$

The component corresponding to this comparison is

$$\frac{(T_1 - T_3)^2}{n[(1)^2 + (-1)^2]} = \frac{(-24)^2}{8(2)} = 36.00.$$

The F ratio for this test is

$$F = \frac{36.00}{4.14} = 8.70.$$

The critical value for a .05-level test is $F_{.95}(1,21) = 4.32$. Since the observed value of F exceeds 4.32, the hypothesis that $\tau_1 - \tau_3 = 0$ is rejected.

The sum of squares for methods, $SS_{\text{methods}} = 50.08$, has two degrees of freedom. Since SS_{C_1} and SS_{C_2} are orthogonal, they represent nonoverlapping (additive) parts of SS_{methods} . Hence

$$SS_{\text{methods}} = SS_{C_1} + SS_{C_2},$$

$$50.08 = .06 + 50.02.$$

C_1 and C_2 are not the only pair of orthogonal comparisons that may be formed. For example,

$$C_3 = (1)T_1 + (0)T_2 + (-1)T_3$$

$$C_4 = (-1)T_1 + (2)T_2 + (-1)T_3$$

are also orthogonal. The components corresponding to these comparisons will also sum to SS_{methods} . In practice the comparisons that are constructed are those having some meaning in terms of the experimental variables; whether these comparisons are orthogonal or not makes little or no difference.

The general form of the hypothesis tested by a comparison is

$$H_1: \sum c_j \mu_j = 0, \quad \sum c_j = 0$$

$$H_2: \sum c_j \mu_j \neq 0,$$

where the c_j 's are the coefficients in the comparison. The statistic used in making the test has the general form

$$F = \frac{(\sum c_j T_j)^2}{(n \sum c_j^2)(MS_{\text{error}})}.$$

The numerator of this F ratio has one degree of freedom. The degrees of freedom for the denominator are those of MS_{error} .

When a large number of comparisons is made following a significant over-all F , some of the decisions which reject H_1 may be due to type 1 error. For example, if 5 independent tests are each made at the .05 level, the probability of a type 1 error in one or more of the five decisions is $1 - (.95)^5 = .23$. If 10 independent tests are made, the probability of a type 1 error's occurring in one or more of the decisions is $1 - (.95)^{10} = .40$. In general, if $\alpha = .05$ and a series of m independent tests are made, then the probability of one or more type 1 errors in the series of m tests is obtained by subtracting the last term of the expansion of $(.05 + .95)^m$ from unity. Equivalently, one may add all terms in the expansion except the last one.

Thus, when the number of comparisons is large, the number of decisions that can potentially be wrong owing to type 1 error can be relatively large.

Tukey and Scheffé have developed methods for constructing simultaneous confidence intervals which avoid the pitfall of permitting the type 1 error to become excessively large. These methods as well as others for dealing with multiple comparisons are discussed in Secs. 3.8 and 3.9.

3.7 Use of Orthogonal Components in Tests for Trend

The total variation of the treatments may be subdivided in many different ways. Depending upon the nature of the experimental variables and the purpose of the experiment, some subdivisions may be meaningful and others may not. The meaningful comparisons need not be the orthogonal ones. If the treatments form a series of equal steps along an ordered scale, i.e., increasing dosage, intensity, complexity, time, etc., then treatment variation may be subdivided into linear, quadratic, cubic, etc., trend components through the use of orthogonal polynomials. This method of subdivision provides information about the form of the relationship between the criterion and the steps along the treatment scale.

Many different curves may be found to fit a given set of empirical data. The use of polynomials of varying degree may or may not be the most appropriate choice for the form of the curve. From one point of view this choice simplifies certain of the statistical problems. From another point of view, the use of higher-degree polynomials may be scientifically meaningless. Polynomials are, however, quite useful in showing the general form of relationships. Within a limited range of values, polynomials can be used to approximate curves which are actually exponential or logarithmic in form.

Caution is required in the use of the methods to be described in this section. The nature of the data and the purpose of the experiment must be considered in evaluation of the results.

In regression analysis, a first-degree (linear) equation is used as the first approximation to the relationship between two variables. Second-degree (quadratic), third-degree (cubic), and higher-degree equations are used if the fit of the linear relationship does not prove to be satisfactory. When there are equal numbers of observations in each of the treatment classes, and when the treatment classes form equal steps along an ordered scale, the work of finding the degree of the best fitting curve is simplified by use of comparisons corresponding to these curves. Once the degree of the best fitting polynomial is found, the regression coefficients for this curve are also readily obtained. Extensive tables exist giving the coefficients for the comparisons corresponding to regression equations of varying degree. The most complete set of coefficients is in the work of Anderson and Houseman (1942). The Fisher and Yates tables (1953) also have an adequate set. A numerical example will be used to illustrate the application of individual components of variation to the problem of finding the degree of the best-fitting curve. (Tests for the best-fitting curve are also called tests for trend.)

Suppose that an experimenter is interested in approximating the form of

the relationship between the degree of complexity of a visual display (i.e., an instrument panel containing a series of dials) and the reaction time of subjects in responding to the displacement of one or more of the dials. After deciding upon a definition of the complexity scale, the experimenter constructs six displays representing six equal steps along this scale. The treatments in this experiment correspond to the degree of complexity of the displays. Random samples of 10 subjects are assigned to each of the displays; each sample is observed under only one of the displays. The criterion used in the analysis is the mean reaction time (or some transformation thereof) for each subject to a series of trials which are as comparable as the displays permit.

Table 3.7-1 Numerical Example

Complexity of display	1	2	3	4	5	6	$k = 6; n = 10$ Total
T_j	100	110	120	180	190	210	910 = G
(i) ΣX_j^2	1180	1210	1600	3500	3810	4610	15,910 = $\Sigma \Sigma X^2$
\bar{T}_j	10.0	11.0	12.0	18.0	19.0	21.0	
(1) = $G^2/kn = (910)^2/60$							= 13,801.67
(2) = $\Sigma \Sigma X^2$							= 15,910
(3) = $(\Sigma T_j^2)/n = (100^2 + 110^2 + \dots + 210^2)/10 = 14,910.00$							
(ii)	$SS_{\text{displays}} = (3) - (1) = 1108.33$						
	$SS_{\text{error}} = (2) - (3) = 1000.00$						
	$SS_{\text{total}} = (2) - (1) = 2108.33$						

Summary data for this experiment are presented in Table 3.7-1. Each T_j in this table represents the sum over the 10 criterion scores for scale j . The sum of the squares of the 10 criterion scores under each scale is also given in part i of this table. Steps in the computation of the sums of squares are given in part ii.

Table 3.7-2 Summary of Analysis of Variance

Source of variation	SS	df	MS	F
Displays	1108.33	5	221.67	11.97**
Experimental error	1000.00	54	18.52	
Total	2108.33	59		

$$** F_{.99}(5,54) = 3.38$$

The analysis of variance appears in Table 3.7-2. Since there are six displays, the degrees of freedom for the displays are $6 - 1 = 5$. Ten subjects were used under each display. Hence the sum of squares within a

single display has degrees of freedom equal to 9; the pooled within-display variation has $6 \times 9 = 54$ degrees of freedom. The pooled within-display variation defines the experimental error. To test the hypothesis that the mean reaction times (or the means of the transformed measures) for the displays are equal,

$$F = \frac{MS_{\text{displays}}}{MS_{\text{error}}} = \frac{221.67}{18.52} = 11.97.$$

The critical value for this test, using the .01 level of significance, is

$$F_{.99}(5, 54) = 3.38.$$

Clearly the data indicate that reaction times for the displays differ. Inspection of the T_j 's in Table 3.7-1 shows that the greater the degree of complexity

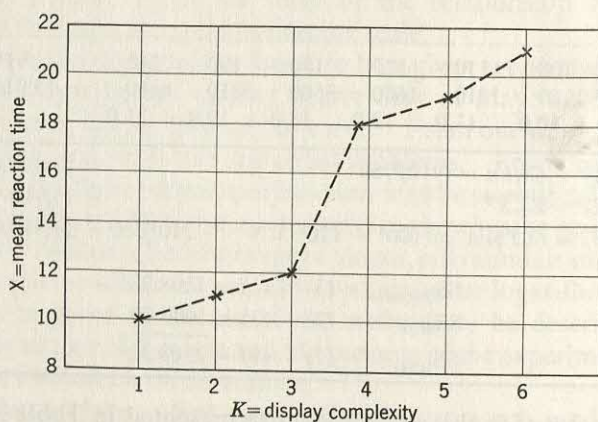


Figure 3.1

the slower the reaction time. The means corresponding to the T_j 's are plotted in Fig. 3.1. This figure shows a strong linear relationship between display complexity and reaction time. There is also the suggestion that an S-shaped curve might be the best-fitting curve; an S-shaped curve corresponds to an equation having degree 3 (cubic).

In the upper left part of Table 3.7-3 are the coefficients for linear, quadratic, and cubic comparisons for $k = 6$. These coefficients were obtained from Table B.10. Notice that the sum of the linear coefficients is zero, that each coefficient differs by 2 units from its neighbor, and that only once in the sequence from -5 to $+5$ does the sign change. For the coefficients defining the quadratic comparison, the signs change twice in the sequence from -5 to $+5$; in the cubic comparison there are three changes of sign in the sequence from -5 to $+5$. The number of times the signs change determines the degree of the polynomial.

The numerical value of the linear comparison for these data is

$$\begin{aligned} C_{\text{lin}} &= (-5)(100) + (-3)(110) + (-1)(120) \\ &\quad + (1)(180) + (3)(190) + (5)(210) \\ &= 850. \end{aligned}$$

Since $D = n\Sigma c^2$, the component of variation corresponding to this comparison has the form

$$\begin{aligned} \frac{C_{\text{lin}}^2}{D_{\text{lin}}} &= \frac{850^2}{10[(-5)^2 + (-3)^2 + (-1)^2 + 1^2 + 3^2 + 5^2]} \\ &= \frac{850^2}{10(70)} = 1032.14. \end{aligned}$$

Thus the variation due to the linear trend in the data is 1032.14 units. From Table 3.7-2 it is seen that the total variation in the displays is 1108.33 units.

Table 3.7-3 Tests for Trend

	T_j :	(1) 100	(2) 110	(3) 120	(4) 180	(5) 190	(6) 210	Σc^2	C	$D = n\Sigma c^2$	C^2/D
Linear		-5	-3	-1	1	3	5	70	850	700	1032.14
(i) Quadratic		5	-1	-4	-4	-1	5	84	50	840	29.76
Cubic		-5	7	4	-4	-7	5	180	-250	1800	34.72
											1096.62

Test for linear trend: $F = \frac{1032.14}{18.52} = 55.73^{**}$

(ii) Test for quadratic trend: $F = \frac{29.76}{18.52} = 1.61$

Test for cubic trend: $F = \frac{34.72}{18.52} = 1.87$

$** F_{.99}(1,54) = 7.14$

Hence $1032.14/1108.33 = 93$ per cent of the variation in the reaction time for the displays may be predicted from a linear regression equation. The test for linear trend is given by

$$F = \frac{\text{linear component}}{\text{MS}_{\text{error}}} = \frac{1032.14}{18.52} = 55.73.$$

The critical value for a .01-level test is $F_{.99}(1, f)$, where f is the degrees of freedom for MS_{error} . In this case $f = 54$, and the critical value is 7.14. Clearly the linear trend is statistically significant. However,

$$1108.33 - 1032.14 = 76.19$$

units of variation are not predicted by the linear regression equation. There is still the possibility that the quadratic or cubic trend might be statistically significant.

An over-all measure of deviations from linearity is given by

$$SS_{\text{nonlin}} = SS_{\text{displays}} - SS_{\text{lin}} = 1108.33 - 1032.14 = 76.19.$$

The degrees of freedom for this source of variation are $k - 2$. An over-all test for nonlinearity is given by

$$F = \frac{SS_{\text{nonlin}}/(k - 2)}{MS_{\text{error}}} = \frac{76.19/4}{18.52} = \frac{19.05}{18.52} = 1.03.$$

For a .01-level test, the critical value is $F_{.99}(4, 54) = 3.68$. Since the F ratio does not exceed this critical value, the data do not indicate that there are significant deviations from linearity. However, when the degrees of freedom for MS_{nonlin} are large, the over-all test may mask a significant higher-order component. In spite of the fact that the over-all test indicates no significant deviations from linearity, if there is a priori evidence that a higher-order component might be meaningful, tests may continue beyond the linear component. Caution is required in the interpretation of significant higher-order trend components when the over-all test for nonlinearity indicates no significant nonlinear trend.

The quadratic comparison is orthogonal to the linear comparison, since the sum of the products of corresponding coefficients is zero, i.e.,

$$(-5)(5) + (-3)(-1) + (-1)(-4) + (1)(-4) + (3)(-1) + (5)(5) = 0.$$

Therefore the quadratic component of variation is part of the 76.19 units of variation which is not predicted by the linear component. The computation of this component is similar to the computation of the linear component, the quadratic coefficients replacing the linear coefficients in the calculation of C and D . The numerical value of the quadratic component is 29.76. This is the increase in predictability that would accrue for the sample data by using a second-degree instead of a first-degree equation. A test on whether or not this increase in predictability is significantly greater than zero uses the statistic

$$F = \frac{\text{quadratic component}}{MS_{\text{error}}} = \frac{29.76}{18.52} = 1.61.$$

The critical value for a .01-level test is $F_{.99}(1, f)$; in this case the critical value is 7.14. The F statistic for quadratic trend is 1.61. Hence the increase in predictability due to the quadratic component is not significantly different from zero.

It is readily verified that the cubic comparison is orthogonal to both the linear and quadratic components. Thus, the cubic component is part of the $1108.33 - 1032.14 - 29.76 = 46.43$ units of variation not predicted by the linear or quadratic trends. The component of variation corresponding to the cubic trend is 34.72. The F statistic for this component is 1.87; the critical value for a .01-level test is 7.14. Thus, the sample data indicate that the cubic component does not increase the predictability by an amount which is significantly different from zero.

Of the total of 1108.33 units of variation due to differences in complexity of the displays, 1096.62 units are predictable from the linear, quadratic, and cubic components. The remaining variation is due to higher-order components, none of which would be significantly different from zero. In summary, the linear trend appears to be the only trend that is significantly greater than zero. Hence a first-degree equation (linear equation) is the form of the best-fitting curve.

In this case the linear equation will have the form

$$X = bK + a,$$

where X = predicted reaction time,
 K = degree of complexity of display,
 b, a = regression coefficients.

Since the degree of complexity of the displays is assumed to be equally spaced along a complexity dimension, the degree of complexity may conveniently be indicated by the integers 1, 2, . . . , 6. The sum of squares for the complexity variable is given by

$$SS_K = \frac{n(k^3 - k)}{12},$$

where k is the number of displays and n is the number of observations under each display. The entry 12 is constant for all values of n and k . The regression coefficient b is given by the relation

$$b = \sqrt{\frac{\text{linear component}}{SS_K}} = \sqrt{\frac{SS_{lin}}{SS_K}}.$$

Equivalently,

$$b = \frac{\lambda_1 C_{lin}}{D_{lin}},$$

where λ_1 is a constant which depends upon k ; tables of coefficients for the orthogonal polynomials will give values of λ_1 .

The symbol SS_{lin} is generally used in regression analysis to indicate the variation predictable from the linear equation. The numerical value of the regression coefficient a is given by the relation

$$a = \bar{X} - b\bar{K}.$$

Computation of the coefficients for the regression equation is summarized in Table 3.7-4.

The linear correlation between degree of complexity and reaction time is given by

$$r = \sqrt{\frac{SS_{lin}}{SS_{total}}} = \sqrt{\frac{1032.11}{2108.33}} = .70.$$

The numerical value for SS_{total} is obtained from Table 3.7-2. The actual fit of the regression equation to the points in Fig. 3.1 is shown in Fig. 3.2.

Table 3.7-4 Computation of Regression Equation

X = reaction time	K = display complexity
$\bar{X} = G/kn = 15.17$	$\bar{K} = (k + 1)/2 = 3.50$

$$SS_K = \frac{n(k^3 - k)}{12} = \frac{(10)(216 - 6)}{12} = 175$$

$$b = \sqrt{\frac{SS_{\text{lin}}}{SS_K}} = \sqrt{\frac{1032.14}{175}} = 2.43$$

$$a = \bar{X} - b\bar{K} = 15.17 - (2.43)(3.50) = 6.67$$

Regression equation: $X = bK + a = 2.43K + 6.67$

The correlation associated with the cubic relationship is -

$$r = \sqrt{\frac{SS_{\text{lin}} + SS_{\text{quad}} + SS_{\text{cubic}}}{SS_{\text{total}}}}$$

$$= \sqrt{\frac{1032.14 + 29.76 + 34.72}{2108.33}} = .72.$$

The tests that have been made indicate that the cubic correlation does not differ significantly from the linear correlation.

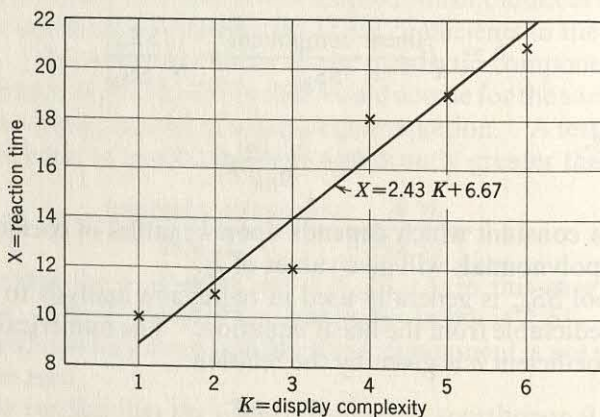


Figure 3.2

In working with higher-order regression equations, in predicting V from U it is convenient to work with the equation in the second of the following forms:

$$V = a_0 + a_1U + a_2U^2 + a_3U^3$$

$$= A'_0\xi'_0 + A'_1\xi'_1 + A'_2\xi'_2 + A'_3\xi'_3$$

In this latter form,

$$A'_0 = \bar{V}, \quad \xi'_0 = 1,$$

$$A'_1 = \frac{C_{\text{lin}}}{D_{\text{lin}}}, \quad \xi'_1 = \lambda_1(U - \bar{U}).$$

$$A'_2 = \frac{C_{\text{quad}}}{D_{\text{quad}}}, \quad \xi'_2 = \lambda_2 \left[(U - \bar{U})^2 - \frac{k^2 - 1}{12} \right],$$

$$A'_3 = \frac{C_{\text{cubic}}}{D_{\text{cubic}}}, \quad \xi'_3 = \lambda_3 \left[(U - \bar{U})^3 - (U - \bar{U}) \frac{3k^2 - 7}{20} \right].$$

The numerical values of the λ_j 's depend upon the value of k , the number of treatments. Numerical values for λ_j will be found in tables of the coefficients.

3.8 Use of the Studentized Range Statistic

Although the hypothesis $\tau_1 = \tau_2 = \dots = \tau_k$ is generally tested by means of an F ratio, there are other methods for testing this same hypothesis. One method is through use of the studentized range statistic, defined by

$$(1) \quad q = \frac{\bar{T}_{\text{largest}} - \bar{T}_{\text{smallest}}}{\sqrt{\text{MS}_{\text{error}}/n}},$$

where n is the number of observations in each \bar{T} . In words, the studentized range statistic is the difference between the largest and smallest treatment means (the range of the treatment means) divided by the square root of the quantity mean-square experimental error over n . A numerically equivalent form of the q statistic is given by

$$(1a) \quad q = \frac{T_{\text{largest}} - T_{\text{smallest}}}{\sqrt{n \text{MS}_{\text{error}}}},$$

where a T represents a sum of n observations. Computationally it is more convenient to compute first the following statistic:

$$(2) \quad F_{\text{range}} = \frac{(T_{\text{largest}} - T_{\text{smallest}})^2}{2n \text{MS}_{\text{error}}}.$$

Then q may readily be shown to be equal to

$$(3) \quad q = \sqrt{2F_{\text{range}}}.$$

Under the hypothesis that $\tau_1 = \tau_2 = \dots = \tau_k$ and under all the assumptions underlying either model I or model II, the sampling distribution of the q statistic is approximated by the studentized range distribution having parameters k = number of treatments and f = degrees of freedom for MS_{error} . The symbol $q_{.99}(k, f)$ designates the 99th percentile point on the

q distribution. Tables of the latter are given in Table B.4. In contrast to the over-all F ratio, which uses all the k treatment means to obtain the numerator, the q statistic uses only the two most extreme means in the set of k . Thus, the q statistic uses less of the information from the experiment than does the F statistic. Use of the F statistic generally leads to a more powerful test with respect to a broader class of alternative hypotheses than does the use of the q statistic. However, there are some alternative hypotheses for which the q statistic leads to a more powerful test. Whether or not the two tests lead to the same decision with respect to the hypothesis being tested depends upon the distributions in the populations from which the

Table 3.8-1 Numerical Example

Treatments . . .	a	b	c	d	e	$k = 5; n = 4$
T_j	10	14	18	14	14	
\bar{T}_j	2.50	3.50	4.50	3.50	3.50	
Source			SS	df	MS	F
Treatments			8.00	4	2.00	4.00
Experiment error			7.50	15	0.50	
Total			15.50	19		

$$F_{.99}(4,15) = 4.89$$

means were obtained. If the distributions are each $N(\mu, \sigma)$, the use of either the F or the q statistic will lead to the same decision. In other cases the decisions reached may differ.

The numerical example summarized in Table 3.8-1 provides an illustration of a case in which the two tests lead to different decisions. Each T_j in this table represents the sum over 4 observations ($n = 4$) under each of 5 treatments ($k = 5$). Since the observed over-all F ratio (4.00) does not exceed the critical value (4.89) for a .01-level test, the data do not contradict the hypothesis that the treatment effects are all equal. The q statistic as computed from (1) is

$$q = \frac{4.50 - 2.50}{\sqrt{.50/4}} = 5.67.$$

The critical value for a .01-level test is $q_{.99}(5,15) = 5.56$. Since the observed q statistic exceeds the critical value, the data contradict the hypothesis that the treatment effects are all equal. In this case the F statistic and the q statistic lead to conflicting decisions. This is so because the means for treatments b , d , and e fall at a single point.

In contrast to the example in Table 3.8-1, consider the example in Table 3.8-2. In this latter table the means tend to be more evenly distributed between the highest and lowest values. SS_{treat} for this case is larger than it is for the case in which the means concentrate at a single point within the range. The experimental error in this example is numerically equal to that in Table 3.8-1. The over-all F value exceeds the critical value for a .01-level test. Since the ranges are the same in the two examples and since the error is also the same, there is no numerical change in the value of the q statistic. If, in practice, the over-all F leads to nonrejection of the hypothesis being tested but the q test leads to rejection of the hypothesis being

Table 3.8-2 Numerical Example

Treatments . . .	<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	$k = 5; n = 4$
T_j	10	12	18	16	14	
\bar{T}_j	2.50	3.00	4.50	4.00	3.50	
Source			SS	df	MS	<i>F</i>
Treatments			10.00	4	2.50	5.00**
Experimental error			7.50	15	0.50	
Total			17.50	19		

$$** F_{.99}(4,15) = 4.89$$

tested, the experimenter should examine his data quite carefully before attempting any interpretation. In most cases, the F test is more powerful than the corresponding q test.

A numerical example will be used to illustrate the marked difference in sampling distributions for statistics constructed in a way which takes into account the order of magnitude of the observed outcomes. Consider a population having the following five elements:

Element	X
1	10
2	15
3	20
4	25
5	30

The collection of all possible samples of size $n = 3$, the sequence within the sample being disregarded, that can be constructed by sampling without replacement after each draw is enumerated in part i of Table 3.8-3.

The statistic d_3 is defined to be the difference between the largest and the smallest of the sample values. The statistic d_2 is the difference between the next largest and the smallest. Under a sampling plan which gives each of

the samples in part i an equal chance of being drawn, the sampling distributions for d_2 and d_3 are the cumulative relative-frequency distributions in part ii. From the latter, one obtains the following probability statements:

$$P(d_2 > 10) = .10,$$

$$P(d_3 > 10) = .70.$$

In words, the probability of large values for d_3 is larger than the corresponding probability for d_2 .

Table 3.8-3 Sampling Distributions for Ordered Differences

	Sample			Sample		
	d_2	d_3		d_2	d_3	
(i)	10, 15, 20	5	10	10, 25, 30	15	20
	10, 15, 25	5	15	15, 20, 25	5	10
	10, 15, 30	5	20	15, 20, 30	5	15
	10, 20, 25	10	15	15, 25, 30	10	15
	10, 20, 30	10	20	20, 25, 30	5	10
(ii)	d_2	Cumulative rel. frequency	d_3	Cumulative rel. frequency		
	20	1.00	20	1.00		
	15	1.00	15	.70		
	10	.90	10	.30		
	5	.60	5	.00		

The principle that emerges is this: When one attempts to draw inferences from statistics determined by taking into account the order of magnitude of the observations within an experiment, the appropriate sampling distribution depends upon the number of ordered steps between the observations from which the statistic was computed.

A modified q statistic is particularly useful in probing the nature of the differences between treatment means following a significant over-all F . The procedure to be described is known as the Newman-Keuls method. For this purpose, the q_r statistic will be used, where r is the number of steps two means (or totals) are apart on an ordered scale. Consider the following treatment totals arranged in increasing order of magnitude:

Order . . .	1	2	3	4	5	6	7
T_j	$T_{(1)}$	$T_{(2)}$	$T_{(3)}$	$T_{(4)}$	$T_{(5)}$	$T_{(6)}$	$T_{(7)}$

In this notation $T_{(1)}$ designates the smallest treatment total and $T_{(7)}$ the largest treatment total. $T_{(7)}$ is defined as being seven steps from $T_{(1)}$, $T_{(6)}$ is six steps from $T_{(1)}$, etc.; $T_{(7)}$ is six steps from $T_{(2)}$, $T_{(6)}$ is five steps from $T_{(2)}$, etc. In general, the number of steps between the ordered totals $T_{(j)}$ and $T_{(i)}$ is $j - i + 1$. Thus,

$$q_7 = \frac{T_{(7)} - T_{(1)}}{\sqrt{nMS_{\text{error}}}};$$

there is only one q_7 statistic. There are, however, two q_6 statistics,

$$q_6 = \frac{T_{(7)} - T_{(2)}}{\sqrt{n\text{MS}_{\text{error}}}} \quad \text{and} \quad q_6 = \frac{T_{(6)} - T_{(1)}}{\sqrt{n\text{MS}_{\text{error}}}}.$$

q_7 corresponds to the ordinary studentized range statistic when $k = 7$. q_r , where r is less than k , is a modified or truncated studentized range statistic. Critical values for q_r are obtained from tables of the studentized range statistic by setting r equal to the range. For example, the critical value for q_6 in a .01-level test is $q_{.99}(6, f)$; the corresponding critical value for q_4 is $q_{.99}(4, f)$, where f is the degrees of freedom for MS_{error} . (Tables of the q statistic are given in Table B.4.)

Rather than working directly with the critical value for the q_r statistic, it is more convenient, in making a large number of tests, to obtain a critical value for the difference between two totals. For example,

$$T_{(7)} - T_{(2)} = q_6 \sqrt{n\text{MS}_{\text{error}}}.$$

Since the critical value for q_6 is $q_{1-\alpha}(6, f)$, the critical value for $T_{(7)} - T_{(2)}$ will be $q_{1-\alpha}(6, f) \sqrt{n\text{MS}_{\text{error}}}$. In general, the critical value for the difference between two treatment totals which are r steps apart on an ordered scale will be $q_{1-\alpha}(r, f) \sqrt{n\text{MS}_{\text{error}}}$.

The use of the q_r statistic in testing the difference between all pairs of means following a significant over-all F will be illustrated by the numerical example in Table 3.8-4. The summary of the analysis of variance appears in part i. Since the observed F ratio (9.40) exceeds the critical value for a .01-level test, the hypothesis that the effects of the seven treatments are all equal is rejected. The seven treatments are designated by the symbols a through g . The totals of the five observations under each of the treatments are arranged in increasing order of magnitude in part ii of this table. For example, treatment c has the smallest total; treatment f has the largest total. A table of differences between the treatment totals also appears in part ii. For example, the entry in column a , row c , is $T_a - T_c = 12 - 10 = 2$; the entry in column g , row a , is $T_g - T_a = 22 - 12 = 10$. In general, the entry in column j , row i , is $T_j - T_i$. Only the entries shown in this table need be computed.

The critical values for the q_r statistic (when $\alpha = .01$) are given in part iii of Table 3.8-4. Since mean square for experimental error has 28 degrees of freedom, $f = 28$. The entry 3.91 is obtained from tables of the studentized range statistic in which $k = 2$ and $f = 28$. The entry 4.48 is obtained from the same row of the tables, but from the column in which $k = 3$. Similarly the entry 4.84 is obtained from the column in which $k = 4$. These entries are critical values for q_r ; to obtain the critical value for the difference between treatment totals which are r steps apart, the critical values are multiplied by

$$\sqrt{n\text{MS}_{\text{error}}} = \sqrt{5(.80)} = \sqrt{4.00} = 2.00.$$

the difference between them must be before this difference exceeds its critical value. Thus, if one examines the data obtained from an experiment and decides to test the difference between the largest and the smallest mean in a set of k means, the critical value for a test of this kind is larger than the critical value for two means which are adjacent to each other on an ordered scale. The larger critical value is required because the sampling distribution of the difference between the largest and smallest means in a set of k will have a greater relative frequency of more extreme values than will the sampling distribution of two adjacent means in a set of k .

There is a prescribed sequence in which tests on the differences between treatment totals in part ii must be made.

1. The first test is on the difference in the upper right-hand corner. This difference is $T_{(7)} - T_{(1)} = T_f - T_c = 15$; the totals here are $r = 7$ steps apart. The critical value for a .01-level test is 10.90. Since the observed difference exceeds 10.90, the hypothesis that $\tau_f = \tau_c$ is rejected. Two asterisks appear in cell (c, f) in part iv to indicate that this hypothesis is rejected. If this hypothesis is not rejected, no additional tests are made.

2. Tests are now made successively on the entries in row c , proceeding from right to left until the first nonsignificant difference is found. The test on the difference $T_{(6)} - T_{(1)} = T_e - T_c = 14$ has the critical value 10.56, since these totals are $r = 6$ steps apart. The test on the difference

$$T_{(5)} - T_{(1)} = T_d - T_c = 12$$

has the critical value 10.18, since these totals are $r = 5$ steps apart. The difference $T_{(4)} - T_{(1)} = T_b - T_c = 8$ has the critical value 9.68, since these totals are $r = 4$ steps apart. This last difference does not exceed its critical value. Hence no additional tests are made in row c . Since the last significant entry occurs in column g , no additional tests made will go beyond column g .

3. The next test is made on the extreme right-hand entry in row a . This entry is $T_{(7)} - T_{(2)} = T_f - T_a = 12$. Because these totals are $r = 6$ steps apart, the critical value is 10.56. Thus, the hypothesis that $\tau_f = \tau_a$ is rejected. If this hypothesis had not been rejected, no additional tests would be made.

4. Tests are continued from right to left in row a either until a nonsignificant entry is obtained or until column g is reached, whichever occurs first. The entry $T_{(6)} - T_{(2)} = T_e - T_a = 12$ has the critical value 10.18, since these totals are $r = 5$ steps apart. The entry $T_{(5)} - T_{(2)} = T_d - T_a = 10$ has the critical value 9.68, since these totals are $r = 4$ steps apart. Since no tests are made beyond column g , no additional tests are made in row a .

5. Tests now proceed from right to left in row d . Tests in this row continue until either a nonsignificant entry is reached or the point at which tests were stopped in the preceding row is reached, whichever occurs first.

The first entry at the right in row d is $T_{(7)} - T_{(3)} = T_f - T_d = 12$; this entry has the critical value 10.18. The next entry in row d ,

$$T_{(6)} - T_{(3)} = T_e - T_d = 11,$$

has the critical value 9.68, since the totals are $r = 4$ steps apart. The next entry, $T_{(5)} - T_{(3)} = T_g - T_d = 9$, has the critical value 8.96, since these totals are $r = 3$ steps apart. Since no tests are made beyond column g , this is the last test made in row d .

6. Tests now proceed from right to left in row b until either a nonsignificant entry occurs or the column at which tests were terminated in the preceding row is reached, whichever occurs first. The first entry on the right in row b , $T_{(7)} - T_{(4)} = T_f - T_b = 7$, has a critical value of 9.68, since the totals are $r = 4$ steps apart. The hypothesis that $\tau_f = \tau_b$ cannot be rejected. No additional tests are made in row b or in any row below it.

The differences which proved to be significant at the .01 level are indicated by asterisks in part iv of Table 3.8-3. The information in part iv may be summarized schematically as follows:

$$\underline{c \ a \ d \ b \ g \ e \ f}.$$

Treatments underlined by a common line do not differ from each other; treatments not underlined by a common line do differ. Thus, treatment f differs from treatments c , a , and d , but treatment f does not differ from treatments e , g , and b . Similarly treatment e differs from c , a , and d but does not differ from b , g , f .

This sequence for making the tests prevents one from arriving at contradictory decisions of the following type: Suppose there are four means in an ordered set and that the difference $T_{(4)} - T_{(1)}$ is close to being significant but does not quite "make it." Further suppose that the difference $T_{(3)} - T_{(1)}$ is just larger than the appropriate critical value. In this case one might be tempted to conclude that there is no significant difference between the largest and the smallest means in the set but that there is a significant difference between the next to the largest and the smallest. Geometrically this conclusion would be equivalent to inferring that the distance between the largest and the smallest of four means is zero but that the distance from the smallest to the next to the largest is greater than zero. Yet the latter distance has to be smaller than the former. Clearly this kind of inference leads to a contradiction.

In general, if the sequence indicated above is followed in making tests on all possible pairs of ordered means, the patterns of significant differences indicated below will *not* occur:

(1)	(2)	(3)	(4)	(5)
	*	—	*	*
		—	*	*
			*	—
			—	—

(1)	(2)	(3)	(4)	(5)
	*	*	—	*
		—	*	—
		*	—	*
			*	—

That is, between any two asterisks in the same row or column there can be no gaps (nonsignificant differences). Further, if the extreme position at the right of a row is a gap, then there can be no asterisks in that row or any row below that row.

It is of interest to compare the critical values for tests on ordered means with tests on individual components made through use of the F statistic. Since $q = \sqrt{2F}$, the critical value for a test on individual components for the example in Table 3.8-3 would be (in comparable units)

$$\sqrt{2F_{.99}(1,28)} = \sqrt{2(7.64)} = 3.91.$$

This is the critical value for totals which are two steps apart. Hence use of the F statistic is equivalent to using the 3.91 as the critical value for all q 's, or using the value 7.82 as the critical value for differences between treatment totals. The critical values for the Newman-Keuls procedure range from 7.82 to 10.90. Hence the use of the F statistic leads to more significant results than does the use of the q_r statistic.

If the meaningful comparisons are relatively few in number and are planned before the data are obtained, the F test associated with individual components of variation should be used. This type of comparison is called an a priori comparison in contrast to comparisons made after inspection of the experimental data; the latter are called a posteriori or post-mortem comparisons. The a priori type is always justified whether or not the overall F is significant. If the k treatments fall into one or more natural groupings in terms of the treatments, tests on ordered differences may be made separately within each of the groupings. Other procedures for making a posteriori comparisons are discussed in the next section.

In contrast with the procedures to be described in the next section, the level of significance for the Newman-Keuls procedure is considered individually with respect to each test. Tests in the next section have a different orientation with respect to the level of significance.

3.9 Alternative Procedures for Making A Posteriori Tests

The Newman-Keuls procedure, described in the last section, keeps the level of significance equal to α for all ordered pairs, no matter how many steps apart the means may be. However, the level of significance with respect to the collection of all tests made, considered as a single test, is considerably lower than α . Thus, the power of the collection of all tests made is less than that associated with an ordinary α -level test. Duncan (1955) has developed a procedure which uses a *protection level* of α for the collection of tests, rather than an α level for the individual tests.

[Scheffé (1959, p. 78) takes issue with the principles underlying the development of the sampling distributions which Duncan uses in obtaining the critical values for tests.]

In terms of individual tests, a protection level with $\alpha = .01$ provides a level of significance equal to .01 for means which differ by two ordered steps; for means differing by three steps, the level of significance is equal to $1 - (.99)^2 = .02$; for means differing by four steps, the level of significance is equal to $1 - (.99)^3 = .03$. Thus, with protection level equal to α , the level of significance for individual tests is numerically higher than α when the pairs are more than two steps apart.

The statistic used in the Duncan procedure is the same as that used in the Newman-Keuls test, namely, q_r , where r is the number of steps apart two means or totals are in an ordered sequence. The steps followed in using the Duncan procedure are identical to those followed in the Newman-Keuls procedure. However, the critical values for the q_r statistic are obtained from tables prepared by Duncan for use with protection levels.

For the case $k = 7$ and degrees of freedom for experimental error equal to 28, critical values for the .01-level Newman-Keuls tests and the .01-level protection level are as follows:

k	2	3	4	5	6	7
Newman-Keuls . . .	3.91	4.48	4.84	5.09	5.28	5.45
Duncan	3.91	4.08	4.18	4.28	4.34	4.39

When $k = 2$, the two procedures have identical critical values. For values of k larger than 2, the Duncan procedure has the smaller critical value. The larger the value of r , the larger the difference between the critical values for the two procedures. Thus, on the average, a larger difference between two means (or two totals) is required for statistical significance under the Newman-Keuls procedure.

To illustrate the difference in size of the type 1 error associated with individual tests, which disregards the order aspect, the q_r statistic may be transformed into an F statistic by means of the relation

$$q_r = \sqrt{2F} \quad \text{or} \quad F = \frac{q_r^2}{2}.$$

For the case discussed in the preceding paragraph, 5.45 is equivalent to an F of 14.85, and 4.39 is equivalent to an F of 8.63. Relative to the critical value for an individual comparison made by means of an F statistic, the type 1 error associated with the difference between two totals which are seven steps apart is as follows:

	Critical F value	"Actual" α
Individual comparison	7.64	.01
Newman-Keuls	14.85	.0005
Duncan	8.63	.007

In other words, if a .01-level test were made on the difference between two means that are seven steps apart on an ordered scale, assuming that MS_{error} has 28 degrees of freedom, and if the order of the means in the sample were disregarded, the critical value would be $F = 7.64$. If, however, the order were taken into account, the equivalent critical value would be 14.85. With a .01 protection level, the equivalent critical value would be 8.63.

A considerably more conservative procedure in terms of keeping the type I error small is the use of $q_{1-\alpha}(k, f)$ as the critical value for all tests, no matter how many steps apart the means may be. Thus, instead of changing the critical value as a function of the number of steps two means are apart on an ordered scale, the critical value for the maximum number of steps is used for all tests. This approach, suggested by Fisher, has been studied and extended by Tukey and will be called the Tukey (*a*) procedure. [This procedure has also been called the honestly significant difference (hsd) procedure.] Compared with the Newman-Keuls and Duncan approaches, fewer significant differences will be obtained. For the example considered in Table 3.8-4, the critical value for the q_r statistic would be 5.45 for all tests; equivalently, the critical value for the difference between two treatment totals would be 10.90 for all differences. If the Tukey (*a*) test were used with the data in Table 3.8-4, part iv of this table would have the following form.

	<i>c</i>	<i>a</i>	<i>d</i>	<i>b</i>	<i>g</i>	<i>e</i>	<i>f</i>
<i>c</i>					**	**	**
<i>a</i>						**	**
<i>d</i>						**	**

The power of Tukey (*a*) tests is lower than those of the Newman-Keuls and the Duncan procedures. Use of the Tukey (*a*) procedure is not necessarily limited to tests on differences between pairs of means; it may also be used in making comparisons involving three or more means. The Tukey (*a*) procedure has this general property: all tests on differences between pairs have a level of significance which is at most equal to α .

Tukey has also proposed a second procedure, which is a compromise between the Tukey (*a*) tests and the Newman-Keuls procedure. The statistic employed in making what will be called Tukey (*b*) tests is again the q_r statistic. However, the critical value is the average for the corresponding value in the Newman-Keuls tests and the critical value for the Tukey (*a*) tests. In symbols, this critical value is

$$\frac{q_{1-\alpha}(k, f) + q_{1-\alpha}(r, f)}{2},$$

where k is the number of means in the set and r is the number of steps between the two means being compared. For example, if $k = 7$ and

$f = 28$, the critical value for a .01-level test for two means which are four steps apart is

$$\frac{5.45 + 4.84}{2} = 5.14.$$

A procedure for making all possible comparisons, not specifically comparisons involving two means, has been developed by Scheffé. An F statistic corresponding to a component of variation is computed, but the critical value for this component is $(k - 1)F_{1-\alpha}(k - 1, f)$. All tests use this critical value. For example, if $k = 7$, $f = 28$, and $\alpha = .01$, the critical value for Scheffé tests is $6[F_{.99}(6, 28)] = 6(3.53) = 21.18$. In terms of the q statistic this is equivalent to a critical value of 6.51. If Scheffé tests were made on the data in Table 3.8-4, the critical value would be $(6.51) \sqrt{nMS_{\text{error}}} = 13.02$ for all differences between treatment totals. The statistically significant differences using this critical value are as follows.

	<i>c</i>	<i>a</i>	<i>d</i>	<i>b</i>	<i>g</i>	<i>e</i>	<i>f</i>
<i>c</i>						**	**

Thus, the Scheffé method applied to testing differences between all possible pairs is even more conservative with respect to type 1 errors than is the Tukey (α) method.

The Scheffé approach has this optimum property: the type 1 error is at most α for any of the possible comparisons. In the original development, Scheffé was concerned with constructing a set of simultaneous confidence intervals on all the comparisons within a subdivision of the entire experiment. Before a set of simultaneous confidence intervals is considered to be true, each of the separate statements must be true. If any one of the confidence intervals in the set is false, then the confidence statement is considered to be false. The simultaneous-confidence-interval approach can be translated into a procedure for making all possible tests on comparisons. This translation has been outlined above.

The data in Table 3.8-4 will be used to summarize the various methods for making a posteriori tests. For $\alpha = .01$, the critical values for differences between pairs of ordered totals are as follows:

Method	<i>k</i> : 2	3	4	5	6	7
Scheffé	13.02	13.02	13.02	13.02	13.02	13.02
Tukey (<i>a</i>)	10.90	10.90	10.90	10.90	10.90	10.90
Tukey (<i>b</i>)	9.36	9.93	10.29	10.54	10.74	10.90
Newman-Keuls	7.82	8.96	9.68	10.18	10.56	10.90
Duncan	7.82	8.16	8.36	8.56	8.68	8.78
Individual comparisons	7.82	7.82	7.82	7.82	7.82	7.82

The last method is primarily for meaningful comparisons planned prior to inspection of the data; it is included in this summary only for purposes of illustration. The Scheffé method is clearly the most conservative with respect to type 1 error; this method will lead to the smallest number of significant differences. In making tests on differences between all possible pairs of means it will yield too few significant results. The Tukey (α) test will also tend to yield too few significant results. Because the Tukey (α) test is applicable in a relatively broad class of situations, and because it is simple to apply, there is much to recommend the Tukey (α) test for general use in making *a posteriori* tests.

3.10 Comparing All Means with a Control

If one of the k treatments in an experiment represents a control condition, the experimenter is generally interested in comparing each treatment with the control condition, regardless of the outcome of the over-all F . There are $k - 1$ comparisons of this kind. Rather than setting a level of significance equal to α for each of the tests, the experimenter may want to have a level equal to α for the collection of the $k - 1$ decisions, considered as a single decision summarizing the outcomes.

Since each of the tests uses the same information on the control condition and a common estimate of experimental error, the tests are not independent. Dunnett (1955) has derived the sampling distribution for a t statistic appropriate for use when level of significance α is desired for the set of all comparisons between several treatments and a control. The parameters of Dunnett's distribution for the t statistic are:

$$\begin{aligned} k &= \text{number of treatments (including the control),} \\ df &= \text{degrees of freedom for } MS_{\text{error}}. \end{aligned}$$

The approach used by Dunnett is a special case of the problem of handling multiple comparisons by constructing a joint confidence interval on the set of all relevant comparisons.

The numerical example in Table 3.10-1 will be used to illustrate the application of Dunnett's t statistic. In this experiment there are four treatment conditions ($k = 4$); three observations ($n = 3$) are made under each of the treatment conditions. One of the treatments represents a standard manufacturing process; the other three represent different methods for manufacturing the same product. The criterion measure in each case is an index of quality of the manufactured product. In this case the over-all $F = 7.05$ exceeds the critical value for a .05-level test. The t statistic for the difference between method j and the standard method is

$$t = \frac{\bar{T}_j - \bar{T}_0}{\sqrt{2MS_{\text{error}}/n}}.$$

The critical value for the collection of $k - 1$ statistics of this form that may

be computed is obtained from the Dunnett tables given in Table B.6. For the data in Table 3.10-1 the critical value for a two-tailed .05-level test is ± 2.94 [that is, $t_{.975}(8) = 2.94$]. This value is found under the column

Table 3.10-1 Numerical Example

	Methods				
	Standard	I	II	III	
	55	55	55	50	$n = 3; k = 4$
	47	64	49	44	
	48	64	52	41	
T_j	150	183	156	135	$G = 624$ $\Sigma(\Sigma X^2) = 33,002$
$\Sigma(X_j^2)$	7538	11,217	8130	6117	
\bar{T}_j	50	61	52	45	

$$(1) = G^2/kn = (624)^2/12 = 32,448.00$$

$$(2) = \Sigma(\Sigma X^2) = 33,002$$

$$(3) = (\Sigma T_j^2)/n = (150^2 + 183^2 + 156^2 + 135^2)/3 = 32,850.00$$

$$SS_{\text{methods}} = (3) - (1) = 402.00$$

$$SS_{\text{error}} = (2) - (3) = 152.00$$

$$SS_{\text{total}} = (2) - (1) = 554.00$$

Source of variation	SS	df	MS	F
Methods	402.00	3	134.00	7.05
Experimental error	152.00	8	19.00	
Total	554.00	11		

$$F_{.95}(3,8) = 4.07$$

headed 4 and the row corresponding to degrees of freedom equal to 8. For example, in comparing method I with the standard,

$$t = \frac{61 - 50}{\sqrt{2(19)/3}} = \frac{11}{3.56} = 3.09.$$

Since the observed t statistic exceeds the critical value, the hypothesis that the two methods of manufacturing yield products having equal average quality index is rejected. The level of significance for this single test is not .05; it is approximately .02. The critical value of 2.94 is associated with the collection of the three tests that would be made. The other two tests are

$$t = \frac{52 - 50}{3.56} = .56,$$

$$t = \frac{45 - 50}{3.56} = 1.41.$$

These three tests may be summarized in the statement that method I differs significantly from the standard but methods II and III do not. This summary decision has significance level equal to .05. In the long run the summary decision reached by the procedures followed will have the equivalent of a type 1 error equal to, at most, .05.

Rather than working with a series of tests of significance, the experimenter may construct a series of confidence intervals. For this purpose it is convenient to introduce the concept of what Tukey has called an allowance, designated by the symbol A . In this context an allowance is defined by

$$A = t_{1-(\alpha/2)} \sqrt{2MS_{\text{error}}/n}.$$

where $t_{1-(\alpha/2)}$ is a value obtained from the Dunnett tables. For the data in Table 3.10-1, with $\alpha = .05$, $A = 2.94(3.56) = 10.5$. The general form for the lower and upper confidence limits is

$$(\bar{T}_j - \bar{T}_0) \pm A.$$

For the data in Table 3.10-1, 95 per cent confidence limits for the collection of confidence statements on the difference $\mu_j - \mu_0$ are as follows:

For method I:	$(61 - 50) \pm 10.5 =$.5 and 21.5.
For method II:	$(52 - 50) \pm 10.5 =$	-8.5 and 12.5.
For method III:	$(45 - 50) \pm 10.5 =$	-15.5 and 5.5.

The joint (simultaneous) confidence coefficient for the intervals defined by these limits is .95. The confidence interval for method I takes the form

$$.5 \leq \mu_I - \mu_0 \leq 21.5.$$

In words, the difference between the mean quality indices for method I and the standard method is between .5- and 21.5-quality index units. This statement and the additional confidence statements for methods II and III have a joint confidence coefficient of .95.

It is of interest to compare the distribution of Dunnett's t statistic with that of Student's t statistic. For the case in which $k = 2$, the two distributions are identical. For k greater than 2, corresponding critical values in the Dunnett tables are larger. For example, with the degrees of freedom for error equal to 10 and $k = 7$, two-tailed tests with joint significance level .05 are equivalent to individual two-tailed tests at the .01 level. It is also of interest to compare the critical values for this type of comparison with those made by means of the Tukey (q) test. The two critical values may be cast in terms of comparable units of measurement by means of the relationship

$$q = \sqrt{2} t.$$

Corresponding critical values for .05 joint significance levels are as follows:

df for error	k	Tukey (α)	Dunnnett
10	8	5.30	4.68
20	6	4.45	3.97
∞	10	4.47	3.89
10	2	3.15	3.15

For values of k greater than 2, the Tukey (α) values are higher than the corresponding Dunnnett values. Since the Tukey (α) test was designed to be appropriate for all possible differences, of which comparisons of each treatment with the control are a subset, it is reasonable to expect that the Tukey (α) tests would require the larger critical values.

3.11 Tests for Homogeneity of Variance

One of the basic assumptions underlying both models I and II is that the variance due to experimental error within each of the treatment populations be homogeneous, that is, $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = \cdots = \sigma_{\epsilon_k}^2$. Moderate departures from this assumption do not, however, seriously affect the sampling distribution of the resulting F statistic. That is, when the variances in the population are not equal, the F statistic using a pooled variance has approximately the same distribution as the F statistic which takes the differences in the population variances into account. The following examples, taken from Box (1954, p. 299), illustrate the effect of lack of homogeneity of variance. In these examples $n = 5$, and $k = 3$.

		Populations			Probability of F exceeding $F_{.95}$
		1	2	3	
(a)	Variances	1	1	1	.050
(b)	Variances	1	2	3	.058
(c)	Variances	1	1	3	.059

In (a) all variances are equal; hence an F statistic which pools variances has probability of .05 of exceeding $F_{.95}$ when $\tau_1 = \tau_2 = \tau_3$. In (b) the variances have the ratio 1:2:3; that is, the variance for the second population is twice the variance of the first population, and the variance of the third population is three times the variance of the first population. For this case, the exact sampling distribution (assuming $\tau_1 = \tau_2 = \tau_3$) for the F statistic shows probability equal to .058 of exceeding $F_{.95}$ obtained from the F statistic which assumes $\sigma_{\epsilon_1}^2 = \sigma_{\epsilon_2}^2 = \sigma_{\epsilon_3}^2$. Using the ordinary F test when the population variances are in the ratio 1:2:3 gives a test having a small positive bias, since relatively more significant results will be obtained than the exact sampling distribution warrants. In (c) the ratio of the

variances is 1:1:3; the F test in this case would also have a small positive bias, since the probability of exceeding $F_{.95}$ is .059 rather than .050. When the number of observations in each treatment class varies considerably, Box indicates that the bias becomes somewhat larger and the direction of the bias is not always positive. Also, the greater the skewness in the distribution of the population variances, the more bias in the resulting tests.

In cases where the experimenter has no knowledge about the effect of the treatments upon the variance, tests for homogeneity of variances may be appropriate as preliminary tests on the model underlying the analysis. There is no need, however, for a high degree of sensitivity in such tests, because F tests are robust with respect to departures from homogeneity of

Table 3.11-1 Numerical Example
($n = 10$)

	Treatment 1	Treatment 2	Treatment 3	Treatment 4
T_j	140	95	83	220
$\sum X_j^2$	2320	1802	869	6640
T_j^2/n	1960.00	902.50	688.90	4840.00
SS_j	360.00	899.50	180.10	1800.00
s_j^2	40.00	99.94	20.01	200.00
				$\sum SS_j = 3239.60$
				$\sum s_j^2 = 359.95$

variance. The experimenter need be concerned about only relatively large departures from the hypothesis of equal population variances.

A relatively simple, but adequate, test of the hypothesis that

$$\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

is one proposed by Hartley. When n is constant for all the k treatments in an experiment, this hypothesis may be tested by means of the statistic

$$F_{\max} = \frac{\text{largest of } k \text{ treatment variances}}{\text{smallest of } k \text{ treatment variances}}$$

$$= \frac{s_{\text{largest}}^2}{s_{\text{smallest}}^2}$$

Under the hypothesis that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$, the sampling distribution of the F_{\max} statistic (assuming independent random samples from normal populations) has been tabulated by Hartley. This distribution is given in Table B.7. The parameters for this distribution are k , the number of treatments, and $n - 1$, the degrees of freedom for each of the treatment class variances. If the observed F_{\max} is greater than the tabled value associated with an α -level test, then the hypothesis of homogeneity of variance is rejected. This test will be illustrated by use of the data in Table 3.11-1.

In this table $k = 4$, and $n = 10$. SS_j is the variation within treatment class j and is given by

$$SS_j = \sum X_j^2 - \frac{T_j^2}{n}.$$

The variance within treatment class j is given by

$$s_j^2 = \frac{SS_j}{n - 1}.$$

The largest of the within-class variances is 200.00; the smallest is 20.01. Thus, the numerical value of the F_{\max} statistic for these data is

$$F_{\max} = \frac{200.00}{20.01} = 10.0.$$

From the tables of the F_{\max} distribution, $F_{\max, .99}(4, 9) = 9.9$. Since the observed value of the F_{\max} statistic is greater than the critical value for a .01-level test, the hypothesis of homogeneity of variance is rejected. This test, referred to as Hartley's test, is in practice sufficiently sensitive for use as a preliminary test in situations where such a test is in order. When the number of observations in each of the treatment classes is not constant, but the n_j 's are relatively close to being equal, the largest of the sample sizes may be used instead of n in obtaining the degrees of freedom required for use in the Hartley tables. This procedure leads to a slight positive bias in the test, i.e., rejecting H_1 more frequently than should be the case.

Another relatively simple test for homogeneity of variance, developed by Cochran, uses the statistic

$$C = \frac{s_{\text{largest}}^2}{\sum s_j^2}.$$

The parameters of the sampling distribution of this statistic are k , the number of treatments, and $n - 1$, the degrees of freedom for each of the variances. Tables of the 95th and 99th percentile point of the distribution of the C statistic are given in Table B.8. For the data in Table 3.11-1,

$$C = \frac{200.00}{359.95} = .56.$$

For a .01-level test the critical value is $C_{.99}(4, 9) = .57$. The observed value of C is quite close to the critical value but does not exceed it. However, the experimenter should on the basis of this result seriously question the tenability of the hypothesis of homogeneity of variance.

In most situations encountered in practice, the Cochran and Hartley tests will lead to the same decisions. Since the Cochran test uses more of the information in the sample data, it is generally somewhat more sensitive than is the Hartley test. In cases where n_j , the number of observations in each treatment class, is not constant but is relatively close, the largest of the n_j 's

may be used in place of n in determining the degrees of freedom needed to enter the tables.

Bartlett's test for homogeneity of variance is perhaps the most widely used test. The routine use of Bartlett's test as a preliminary test on the model underlying the analysis of variance is not, however, recommended. Only in relatively few cases is Bartlett's test useful. From the computational point of view it is more complex than is either the Hartley test or the Cochran test. In Bartlett's test the n_j 's in each of the treatment classes need not be equal; however, no n_j should be smaller than 3, and most n_j 's should be larger than 5. The statistic used in Bartlett's test is

$$\chi^2 = \frac{2.303}{c} (f \log \text{MS}_{\text{error}} - \sum f_j \log s_j^2),$$

where $f_j = n_j - 1 = \text{degrees of freedom for } s_j^2$,

$f = \sum f_j = \text{degrees of freedom for MS}_{\text{error}}$,

$$c = 1 + \frac{1}{3(k-1)} \left(\sum \frac{1}{f_j} - \frac{1}{f} \right),$$

$$\text{MS}_{\text{error}} = \frac{\sum SS_j}{\sum f_j}.$$

When $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$, the sampling distribution of the χ^2 statistic is approximated by the χ^2 distribution having $k - 1$ degrees of freedom.

The data in Table 3.11-1 will be used to illustrate the computation of the χ^2 statistic. Computations will be indicated for the case in which the n_j 's are not assumed to be equal. For this case MS_{error} is most readily obtained from

$$\text{MS}_{\text{error}} = \frac{\sum SS_j}{\sum f_j} = \frac{3239.60}{36} = 89.99,$$

since $\sum f_j = 9 + 9 + 9 + 9 = 36$. Other items required for the computation of the χ^2 statistic are

$$f \log \text{MS}_{\text{error}} = 36 \log 89.99 = 36(1.954) = \underline{70.344}$$

$$f_1 \log s_1^2 = 9 \log 40.00 = 9(1.602) = 14.418$$

$$f_2 \log s_2^2 = 9 \log 99.94 = 9(1.999) = 17.991$$

$$f_3 \log s_3^2 = 9 \log 20.01 = 9(1.301) = 11.709$$

$$f_4 \log s_4^2 = 9 \log 200.00 = 9(2.301) = \underline{20.709}$$

$$\sum f_j \log s_j^2 = 64.827$$

$$c = 1 + \frac{1}{3(3)} \left(\frac{1}{9} + \frac{1}{9} + \frac{1}{9} + \frac{1}{9} - \frac{1}{36} \right)$$

$$= 1 + \frac{1}{9} \left(\frac{15}{36} \right)$$

$$= 1.046.$$

From these terms, the χ^2 statistic in Bartlett's test is

$$\chi^2 = \frac{2.303}{1.046} (70.344 - 64.827) = 12.14.$$

The larger the variation between the s_j^2 's, the larger will be the value of the χ^2 statistic. For a .01-level test of the hypothesis that $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ the critical value is $\chi_{.99}^2(3) = 11.3$. Since the observed value of the χ^2 statistic is larger than the critical value, the experimental data do not support the hypothesis being tested.

For the data in Table 3.11-1, the Hartley, Cochran, and Bartlett tests give comparable results. The Hartley test uses what is equivalent to the range of the sample variances as a measure of heterogeneity, whereas the Bartlett test uses what is equivalent to the ratio of the arithmetic mean to the geometric mean of the variances. The sampling distribution of the latter measure has a smaller standard error and hence provides a more powerful test of the hypothesis being tested. For purposes of detecting large departures from the hypothesis of homogeneity of variance, either the Hartley or the Cochran test is adequate in most cases occurring in practice.

There is some evidence that all the tests for homogeneity of variance that have been discussed above are oversensitive to departures from normality of the distributions of the basic observations. Bartlett and Kendall have proposed a test for homogeneity of variance which is less sensitive to non-normality than any of the above tests. A detailed description and applications of the Bartlett and Kendall test will be found in the work of Odeh and Olds (1959).

3.12 Unequal Sample Sizes

The plan of an experiment may call for an equal number of observations under each treatment, but the completed experiment may not meet this objective. For comparable precision in the evaluation of each treatment effect this objective is a highly desirable one, assuming that the variances for the treatment classes are equal. Circumstances not related to the experimental treatments often prevent the experimenter from having an equal number of observations under each treatment. For example, in animal research deaths may occur from causes in no way related to the experimental treatments. In areas of research in which people are the subjects, it may be that only intact groups can be handled; such intact groups may vary in size.

In the earlier sections of this chapter it is generally assumed that random samples of size n were assigned at random to each of the treatments. In this section it will be assumed that a random sample of size n_1 is assigned to treatment 1, a random sample of size n_2 is assigned to treatment 2, etc. The size of the random sample is not assumed to be constant for all treatments. The form of the definitions of the sums of squares is different from

those appropriate for the case in which the sample size is constant. The notation that will be used is outlined in Table 3.12-1. The number of treatments in the experiment is k . The treatments are designated by the symbols $1, 2, \dots, j, \dots, k$, where the symbol j represents any treatment

Table 3.12-1 Notation

	Treatment 1	Treatment 2	...	Treatment j	...	Treatment k	
Number of observations	n_1	n_2		n_j		n_k	$N = \sum n_j$
Sum of observations	T_1	T_2		T_j		T_k	$G = \sum T_j$
Mean of observations	\bar{T}_1	\bar{T}_2		\bar{T}_j		\bar{T}_k	$\bar{G} = G/N$
Sum of squares of observations	$\sum X_1^2$	$\sum X_2^2$		$\sum X_j^2$		$\sum X_k^2$	$\sum(\sum X_j^2)$
T_j^2/n_j	T_1^2/n_1	T_2^2/n_2		T_j^2/n_j		T_k^2/n_k	
Within-class variation	SS_1	SS_2	...	SS_j	...	SS_k	
Within-class variance	$s_1^2 = \frac{SS_1}{n_1 - 1}$	$s_2^2 = \frac{SS_2}{n_2 - 1}$...	$s_j^2 = \frac{SS_j}{n_j - 1}$...	$s_k^2 = \frac{SS_k}{n_k - 1}$	

within the set. The size of the sample observed under treatment j is designated by the symbol n_j . The total number of elements in the experiment is

$$n_1 + n_2 + \dots + n_k = N.$$

Computational formulas are summarized in Table 3.12-2. Symbols (1)

Table 3.12-2 Computational Formulas

(1) $= \frac{G^2}{N}$	(2) $= \sum(\sum X_j^2)$	(3) $= \sum\left(\frac{T_j^2}{n_j}\right)$
$SS_{\text{treat}} = (3) - (1)$		$df_{\text{treat}} = k - 1$
$SS_{\text{error}} = (2) - (3)$		$df_{\text{error}} = N - k$
$SS_{\text{total}} = (2) - (1)$		$df_{\text{total}} = N - 1$

and (2) have the same general form as they do for the case of equal n 's. However, symbol (3) is different; that is,

$$(3) = \frac{T_1^2}{n_1} + \frac{T_2^2}{n_2} + \dots + \frac{T_k^2}{n_k}.$$

Thus each T_j^2 must be divided by its n_j before the summation is made. The degrees of freedom for each of the sums of squares are also shown in Table 3.12-2. For SS_{error} the number of degrees of freedom is the pooled degrees of freedom for the variation within each of the treatments, i.e.,

$$\begin{aligned} df_{\text{error}} &= (n_1 - 1) + (n_2 - 1) + \dots + (n_k - 1) \\ &= N - k. \end{aligned}$$

Table 3.12-3 Numerical Example

Treatment 1		Treatment 2	Treatment 3
3		7	3
2		8	2
4		4	1
3		10	2
1		6	4
5			2
			3
			1
(i)			
Treatment 1		Treatment 2	Treatment 3
3		7	3
2		8	2
4		4	1
3		10	2
1		6	4
5			2
			3
			1
(ii)			
$n_1 = 6$		$n_2 = 5$	$n_3 = 8$
$T_1 = 18$		$T_2 = 35$	$T_3 = 18$
$\Sigma X_1^2 = 64$		$\Sigma X_2^2 = 265$	$\Sigma X_3^2 = 48$
$\frac{T_1^2}{n_1} = 54.00$		$\frac{T_2^2}{n_2} = 245.00$	$\frac{T_3^2}{n_3} = 40.50$
$SS_1 = 10.00$		$SS_2 = 20.00$	$SS_3 = 7.50$
$s_1^2 = 2.00$		$s_2^2 = 5.00$	$s_3^2 = 1.07$
$\bar{T}_1 = 3.00$		$\bar{T}_2 = 7.00$	$\bar{T}_3 = 2.25$
			$n_4 = 7$
			$T_4 = 66$
			$\Sigma X_4^2 = 658$
			$\frac{T_4^2}{n_4} = 622.29$
			$SS_4 = 35.71$
			$s_4^2 = 5.95$
			$\bar{T}_4 = 9.43$
			$N = 26$
			$G = 137$
			$\Sigma(\Sigma X_j^2) = 1035$
			$\Sigma\left(\frac{T_j^2}{n_j}\right) = 961.79$
			$\Sigma SS_j = 73.21$
			$\bar{G} = \frac{137}{26} = 5.27$
(iii)			
(1) = $G^2/N = (137)^2/26 = 721.88$		(2) = $\Sigma \Sigma X^2 = 1035$	
		$SS_{\text{treat}} = (3) - (1) = 239.91$	
		$SS_{\text{error}} = (2) - (3) = 73.21$	
		$SS_{\text{total}} = (2) - (1) = 313.12$	
		(3) = $\Sigma(T_j^2/n_j) = 961.79$	

The computational formulas given in Table 3.12-2 become those for the case of equal sample sizes when n_j is constant for all treatments. Hence the computational formulas for equal n 's are simplified special cases of those given in Table 3.12-2.

The basic partition of the over-all variation now has the form

$$(1) \quad \begin{aligned} \Sigma \Sigma (X_{ij} - \bar{G})^2 &= \Sigma \Sigma (X_{ij} - \bar{T}_j)^2 + \Sigma n_j (\bar{T}_j - \bar{G})^2, \\ SS_{\text{total}} &= SS_{\text{error}} + SS_{\text{treat}}. \end{aligned}$$

Thus in the computation of SS_{treat} each $(\bar{T}_j - \bar{G})^2$ is multiplied by n_j . Hence the variation due to the treatments is a *weighted* sum of the squared deviations of the treatment means about the grand mean; squared deviations based upon a large number of observations are given greater weight than those based upon a small number of observations. (A definition of SS_{treat} which assigns equal weight to the squared deviations is given at the end of this section. The latter definition is to be preferred in cases where differences in the n_j 's have no direct meaning in terms of the population about which inferences are being drawn.)

A numerical example is given in Table 3.12-3. Data in part i represent the basic observations. Detailed summary data are given in part ii. The symbols used in part ii are defined in Table 3.12-1. For example,

$$SS_1 = \Sigma X_1^2 - \frac{T_1^2}{n_1} = 64 - 54.00 = 10.00.$$

As a rough check for homogeneity of variance,

$$F_{\text{max}} = \frac{s_{\text{largest}}^2}{s_{\text{smallest}}^2} = \frac{5.95}{1.07} = 5.56.$$

The largest of the n_j 's is 8. For a .05-level test, an approximate critical value is $F_{\text{max},ss}(k=4, df=8-1=7) = 8.44$. The data do not contradict the hypothesis of homogeneity of variance.

Use of the computational formulas is illustrated in part iii. There are alternative methods for computing these sums of squares. SS_{error} may be obtained from

$$SS_{\text{error}} = \Sigma SS_j = 73.21.$$

From the definition,

$$\begin{aligned} SS_{\text{treat}} &= \Sigma n_j (\bar{T}_j - \bar{G})^2 \\ &= 6(3.00 - 5.27)^2 + 5(7.00 - 5.27)^2 + 8(2.25 - 5.27)^2 + 7(9.43 - 5.27)^2 \\ &= 239.95. \end{aligned}$$

For purposes of showing just what it is that forms the sum of squares for error and treatments, these alternative computational methods are more revealing, but they also involve more computational effort.

The analysis of variance is summarized in Table 3.12-4. A test on the hypothesis that all the treatment effects are equal is given by the F ratio,

$$F = \frac{MS_{\text{treat}}}{MS_{\text{error}}} = \frac{79.97}{3.33} = 24.02.$$

The critical value for a .01-level test is $F_{.99}(3,22) = 4.82$. The data contradict the hypothesis of no differences in treatment effects. Inspection of the treatment means indicates that treatments 2 and 4 are quite different from treatments 1 and 3.

A comparison has the same form as that for the case of equal n 's, that is,

$$C = c_1\bar{T}_1 + c_2\bar{T}_2 + \cdots + c_k\bar{T}_k, \quad \text{where } \sum c_j = 0.$$

Table 3.12-4 Analysis of Variance

Source of variation	SS	df	MS	F
Treatments	239.91	3	79.97	24.02**
Experimental error	73.21	22	3.33	
Total	313.12	25		

$$**F_{.99}(3,22) = 4.82$$

A component of variation corresponding to a comparison is in this case

$$SS_C = \frac{C^2}{(c_1^2/n_1) + (c_2^2/n_2) + \cdots + (c_k^2/n_k)}$$

Two comparisons,

$$C_1 = c_{11}\bar{T}_1 + c_{12}\bar{T}_2 + \cdots + c_{1k}\bar{T}_k,$$

$$C_2 = c_{21}\bar{T}_1 + c_{22}\bar{T}_2 + \cdots + c_{2k}\bar{T}_k,$$

are orthogonal if

$$\frac{c_{11}c_{21}}{n_1} + \frac{c_{12}c_{22}}{n_2} + \cdots + \frac{c_{1k}c_{2k}}{n_k} = 0$$

These definitions reduce to those given for equal n 's when all the n_j 's are equal.

Computational procedures for comparisons will be illustrated through use of the data in Table 3.12-3. The component of variation corresponding to the difference between treatments 2 and 4 is

$$SS_C = \frac{(\bar{T}_2 - \bar{T}_4)^2}{[(1)^2/n_2] + [(-1)^2/n_4]} = \frac{(7.00 - 9.43)^2}{\frac{1}{5} + \frac{1}{7}} = 17.20.$$

To test the hypothesis that $\tau_2 = \tau_4$, the statistic used is

$$F = \frac{SS_C}{MS_{\text{error}}} = \frac{17.20}{3.33} = 5.17.$$

The numerator of this statistic has 1 degree of freedom, the denominator 22 degrees of freedom (the degrees of freedom for MS_{error}). For a .01-level test the critical value is $F_{.99}(1,22) = 7.94$. Since the observed F statistic does not exceed the critical value, the data do not contradict the hypothesis that $\tau_2 = \tau_4$.

As another example, suppose that a meaningful comparison planned in advance of the experiment is that of treatment 4 with all others combined; i.e., does treatment 4 differ from the average of all other treatment effects? This comparison is given by

$$SS_C = \frac{(3\bar{T}_4 - \bar{T}_1 - \bar{T}_2 - \bar{T}_3)^2}{(c_4^2/n_4) + (c_1^2/n_1) + (c_2^2/n_2) + (c_3^2/n_3)} = \frac{[3(9.43) - (3.00) - (7.00) - (2.25)]^2}{\frac{9}{7} + \frac{1}{6} + \frac{1}{5} + \frac{1}{8}} = 144.54.$$

To test the hypothesis that $\tau_4 = (\tau_1 + \tau_2 + \tau_3)/3$, the statistic used is

$$F = \frac{SS_C}{MS_{\text{error}}} = \frac{144.54}{3.33} = 43.40.$$

The critical value for a .01-level test is $F_{.99}(1,22) = 7.94$. Hence the data clearly contradict the hypothesis that the effect of treatment 4 is equal to the average effect of the other three treatments.

When the n_j 's do not differ markedly, the Newman-Keuls, the Duncan, or either of the Tukey methods may be adapted for use in making tests on differences between all pairs of means. The Newman-Keuls method will be used to illustrate the principles involved. With unequal sample sizes it is convenient to work with the treatment means. (For the case of equal sample sizes it is more convenient to work with the treatment totals.) The example in Table 3.12-3 will be used to illustrate the numerical operations. Part i of Table 3.12-5 gives the treatment means arranged in order of increasing magnitude. The differences between all possible pairs of means are shown. For example, the entry 7.18 in the first row is the difference $9.43 - 2.25$. The entry 4.75 is the difference $7.00 - 2.25$. In general an entry in this table is the difference between the mean at the top of the column and the mean at the left of the row.

The statistic to be used in making tests on these differences is q_r ,

$$q_r = \frac{\bar{T}_j - \bar{T}_{j'}}{\sqrt{MS_{\text{error}}/n}},$$

where r is the number of steps the two means are apart on an ordered scale. The n in the expression $\sqrt{MS_{\text{error}}/n}$ refers to the number of observations in each of the means and is assumed to be constant. If the n_j 's do not differ markedly from each other, the harmonic mean of the n_j 's may be used instead of n in this expression. The harmonic mean \tilde{n} is defined as

$$\tilde{n} = \frac{k}{(1/n_1) + (1/n_2) + \cdots + (1/n_k)}.$$

For the numerical example,

$$\tilde{n} = \frac{4}{\frac{1}{6} + \frac{1}{5} + \frac{1}{8} + \frac{1}{7}} = 6.30.$$

Since the degrees of freedom for MS_{error} are 22, the critical values for the q_r statistic are found in the tables of the studentized range statistic in the row corresponding to 22 degrees of freedom. The critical values for a .01-level test are given in part ii of Table 3.12-5. Thus 3.99 is the critical

Table 3.12-5 Tests on Differences between All Pairs of Means

Treatments		3	1	2	4
Means		2.25	3.00	7.00	9.43
(i)	3	2.25	—	0.75	4.75
	1	3.00	—	4.00	6.43
	2	7.00	—	—	2.43
	4	9.43	—	—	—
			$r = 2$	$r = 3$	$r = 4$
(ii)	$q_{.99}(r, 22)$		3.99	4.59	4.96
(iii)	$\sqrt{MS_{\text{error}}/\tilde{n}} q_{.99}(r, 22)$		2.90	3.34	3.61
		3	1	2	4
(iv)	3			**	**
	1			**	**
	2				
	4				

value for the q_r statistic when $r = 2$, that is, when the means are two steps apart; 4.59 is the critical value for q_r when $r = 3$. In making several tests it is convenient to work with the critical value of the difference between a pair of means rather than the critical value of q_r . Since

$$\sqrt{MS_{\text{error}}/\tilde{n}} q_r = T_j - T_{j'},$$

the critical value for the difference between two means is

$$\sqrt{MS_{\text{error}}/\tilde{n}} q_{1-\alpha}(r, \text{df}).$$

The numerical value of $\sqrt{MS_{\text{error}}/\tilde{n}}$ is in this case $\sqrt{3.33/6.30} = .727$. Hence the critical values for .01-level tests on the differences between pairs of means are given by multiplying the entries in part ii of Table 3.12-5 by .727. These values are given in part iii. For example, the entry $2.90 = (.727)(3.99)$.

The sequence in which the tests must be made is given in Sec. 3.8. This sequence must be followed here.

1. The first test made is on the difference 7.18 in the upper right of part i. Since this entry is the difference between two means that are four steps apart, the critical value is 3.61. Hence the hypothesis that $\tau_3 = \tau_4$ is contradicted by the experimental data.

2. The next test is on the entry 4.75, the difference between two means which are three steps apart. The critical value for this test is 3.34. Hence the data contradict the hypothesis that $\tau_2 = \tau_3$.

3. The entry .75, which is the difference between two means that are two steps apart, is tested next. The critical value is 2.90. Hence the data do not contradict the hypothesis that $\tau_1 = \tau_3$.

4. The entry 6.43 is tested against the critical value 3.34, since this entry is the difference between two means which are three steps apart. Hence the data contradict the hypothesis that $\tau_1 = \tau_4$.

5. The entry 4.00 is tested against the critical value 2.90. The data contradict the hypothesis that $\tau_1 = \tau_2$.

6. The entry 2.43 is tested against the critical value 2.90. The data do not contradict the hypothesis that $\tau_2 = \tau_4$.

A summary of the tests is given in part iv. The cells with asterisks indicate that the corresponding differences are statistically significant at the .01 level. Schematically this summary may be represented as follows:

$$\begin{array}{ccc} \underline{3} & \underline{1} & \underline{2} \quad \underline{4} \end{array}$$

Treatments underlined by a common line do not differ; treatments not underlined by a common line do differ. Hence treatments 2 and 4 differ from treatments 3 and 1, but there is no difference between treatments 2 and 4 and no difference between 3 and 1.

In adapting the Duncan method or either of the Tukey methods to the case of unequal sample sizes, the harmonic mean \bar{n} is used in place of n . For example, in the Tukey (α) test the critical value for .01-level tests on all differences would be $\sqrt{MS_{\text{error}}/\bar{n}} q_{.99}(k, df)$, where k is the number of treatments and df is the degrees of freedom for MS_{error} . For the numerical example, this critical value is 3.65. In this case the Tukey (α) test would give outcomes identical with those obtained by means of the Newman-Keuls test.

The definition of SS_{treat} given earlier in this section requires that each $(\bar{T}_j - \bar{G})^2$ be weighted by n_j . If the error variances for each of the treatments are equal, this weighting procedure essentially weights each squared deviation by a factor proportional to its standard error, which is σ_e^2/n_j . If, on the other hand, the n_j 's are in no way related to the hypothesis being tested and it is desired to give each treatment mean a numerically equal weight in determining SS_{treat} , then the latter source of variation may be defined as

$$SS_{\text{treat}} = \bar{n} \sum (\bar{T}_j - \bar{G})^2,$$

where

$$\bar{G} = \frac{\sum \bar{T}_j}{k}.$$

This definition of \bar{G} differs from that used earlier in this section. If the latter definition of SS_{treat} is used, $SS_{\text{treat}} + SS_{\text{error}}$ will not be numerically equal to SS_{total} .

3.13 Determination of Sample Size

The level of significance of an over-all F test in the analysis of variance sets an upper bound on the type I error in the decision rule. By having a suitably large n under each treatment, the power of the test with respect to a specified alternative hypothesis may be made as large as desired. The specification of the alternative hypothesis is in terms of the parameter

$$\phi' = \sqrt{\frac{\sum(\mu_i - \mu)^2/k}{\sigma^2}},$$

where σ^2 is the experimental error per experimental unit.

The charts given in Table B.11 estimate the power of the F test in the analysis of variance as a function of ϕ' , k = number of treatments, n = number of experimental units under each treatment, and α = level of significance of test.

Before these charts may be used, some estimate of ϕ' must be obtained. This estimate need be only a rough one. Past experimentation, a pilot study, familiarity with the subject matter—all these sources of information will contribute to the estimation procedure. To illustrate one method for estimating ϕ' , suppose that the experimenter considers a difference

$$\mu_i - \mu = 5$$

to be a practically important difference. Suppose that $k = 4$ and that σ^2 is estimated to be 100. Then

$$\frac{\sum(\mu_i - \mu)^2}{k} = \frac{5^2 + 5^2 + 5^2 + 5^2}{4} = 25,$$

and

$$\phi' = \sqrt{\frac{25}{100}} = .50.$$

If the over-all test is to be made at the .01 level of significance, and if power .90 is desired with respect to $\phi' = .50$, then, from the chart in Table B.11 in which $k = 4$ and $\alpha = .01$, it will be found that $n = 20$ experimental units per treatment are required, i.e., a total of 80 experimental units for the experiment.

If, on the other hand, the over-all test is to be made at the .05 level of significance and power .90 is required with respect to ϕ' , then $n = 15$ experimental units per treatment will be needed.

CHAPTER 4

Single-factor Experiments Having Repeated Measures on the Same Elements

4.1 Purpose

In experimental work in the behavioral sciences the elements forming the statistical population are frequently people. Because of large differences in experience and background, the responses of people to the same experimental treatment may show relatively large variability. In many cases, much of this variability is due to differences between people existing prior to the experiment. If this latter source of variability can be separated from treatment effects and experimental error, then the sensitivity of the experiment may be increased. If this source of variability cannot be estimated, it remains part of the uncontrolled sources of variability and is thus automatically part of the experimental error.

One of the primary purposes of experiments in which the same subject is observed under each of the treatments is to provide a control on differences between subjects. In this type of experiment, treatment effects for subject i are measured relative to the average response made by subject i on all treatments. In this sense each subject serves as his own control—responses of individual subjects to the treatments are measured in terms of deviations about a point which measures the average responsiveness of that individual subject. Hence variability due to differences in the average responsiveness of the subjects is eliminated from the experimental error (if an additive model is appropriate).

Experiments in which the same elements are used under all the k treatments require k observations on each element. Hence the term *repeated measurements* to describe this kind of design. To the extent that unique characteristics of the individual elements remain constant under the different treatments, pairs of observations on the same elements will tend to be positively correlated. More generally, the observations will be *dependent*

rather than independent. If the population distributions involved are multivariate normal, the terms *dependent* and *correlated* are synonymous; analogously, the terms *independent* and *uncorrelated* are synonymous in this context. Since the models that will be used are assumed to have underlying multivariate normal distributions, correlated measurements imply statistically dependent measurements. The designs in this chapter may be said to involve correlated, or dependent, observations.

The notation to be used and general computational procedures to be followed are given in the next section. The rationale underlying the analysis and special uses of these designs is presented in later sections.

4.2 Notation and Computational Procedures

Notation for this type of design will be illustrated in terms of people as the elements of the statistical population. However, the notation is not

Table 4.2-1 Notation

Person	Treatment						Total	Mean
	1	2	...	j	...	k		
1	X_{11}	X_{12}		X_{1j}		X_{1k}	P_1	\bar{P}_1
2	X_{21}	X_{22}		X_{2j}		X_{2k}	P_2	\bar{P}_2
.
.
i	X_{i1}	X_{i2}		X_{ij}		X_{ik}	P_i	\bar{P}_i
.
.
n	X_{n1}	X_{n2}		X_{nj}		X_{nk}	P_n	\bar{P}_n
Total	T_1	T_2	...	T_j	...	T_k	G	\bar{G}
Mean	\bar{T}_1	\bar{T}_2	...	\bar{T}_j	...	\bar{T}_k		

restricted to this case. In Table 4.2-1 the symbol X_{11} represents the measurement on person 1 under treatment 1, X_{12} the measurement on person 1 under treatment 2, X_{1j} the measurement of person 1 under treatment j . In general the first subscript to an X indicates the person observed and the second subscript the treatment under which the observation is made.

The symbol P_1 represents the sum of the k observations on person 1, P_2 the sum of the k observations on person 2, P_i the sum of the k observations on person i . In summation notation,

$$P_i = \sum_j X_{ij};$$

that is, P_i is the sum of the k entries in row i . Summation over the subscript j is equivalent to summing over all columns within a single row. The

mean of the observations on person i is

$$\bar{P}_i = \frac{P_i}{k}.$$

The symbol T_1 represents the sum of the n observations under treatment 1, T_2 the sum of the n observations under treatment 2, T_j the sum of the n observations under treatment j . In summation notation,

$$T_j = \sum_i X_{ij}.$$

Summation over the subscript i is equivalent to summing all entries in a single column. The mean of the n observations under treatment j , designated \bar{T}_j , is

$$\bar{T}_j = \frac{T_j}{n}.$$

The sum of the kn observations in the experiment, designated G , is

$$G = \sum P_i = \sum T_j = \sum \sum X_{ij}.$$

The symbol $\sum \sum X_{ij}$ represents the sum over all observations in the experiment. The grand mean of all observations, designated \bar{G} , is

$$\bar{G} = \frac{G}{kn} = \frac{\sum \bar{P}_i}{n} = \frac{\sum \bar{T}_j}{k}.$$

In the analysis of this type of experiment, the total variation is divided into two parts: one part is a function of differences between the means of the people; the other part is a function of the pooled variation within individuals. The total variation is

$$(1) \quad SS_{\text{total}} = \sum \sum (X_{ij} - \bar{G})^2,$$

the sum of the squared deviations of each observation about the grand mean. This source of variation has $kn - 1$ degrees of freedom. That part of the total variation due to differences between the means of the people is

$$(2) \quad SS_{\text{between people}} = k \sum (\bar{P}_i - \bar{G})^2.$$

In words, the between-people variation is a function of the squared deviations of the means for the people about the grand mean. Alternatively, this source of variation may be viewed as due to the differences between all possible pairs of \bar{P}_i ; the larger such differences, the larger this source of variation. Since there are n means, this source of variation has $n - 1$ degrees of freedom.

The variation within person i is

$$SS_{\text{w. person } i} = \sum_j (X_{ij} - \bar{P}_i)^2,$$

the sum of the squared deviations of the observations on person i about the mean for person i . This source of variation has $k - 1$ degrees of freedom. The pooled within-person variation, designated $SS_{w. \text{people}}$, is

$$(3) \quad SS_{w. \text{people}} = \sum_i SS_{w. \text{person } i} = \sum_i \sum_j (X_{ij} - \bar{P}_i)^2.$$

Since the variation within each person has $k - 1$ degrees of freedom, the pooled within-person variation will have $n(k - 1)$ degrees of freedom. It is readily shown that the between- and within-people sources of variation are statistically independent and that

$$SS_{\text{total}} = SS_{\text{between people}} + SS_{w. \text{people}}.$$

The degrees of freedom corresponding to these sources of variation are also additive,

$$kn - 1 = (n - 1) + n(k - 1).$$

To show this partition of SS_{total} algebraically, let

$$b_{ij} = X_{ij} - \bar{P}_i,$$

$$a_i = \bar{P}_i - \bar{G}.$$

Then,
$$\sum_j b_{ij} = 0 \text{ for all } i, \quad \sum_i a_i = 0, \quad \sum_j a_i = na_i.$$

Hence,
$$\sum_i \sum_j a_i b_{ij} = \sum_i a_i (\sum_j b_{ij}) = \sum_i (0) = 0.$$

From the definitions of b_{ij} and a_i , it follows that

$$X_{ij} - \bar{G} = b_{ij} + a_i.$$

Hence,
$$\begin{aligned} SS_{\text{total}} &= \sum_i \sum_j (X_{ij} - \bar{G})^2 = \sum_i \sum_j (b_{ij} + a_i)^2 \\ &= \sum_i \sum_j b_{ij}^2 + \sum_i \sum_j a_i^2 + 2 \sum_i \sum_j a_i b_{ij} \\ &= \sum_i \sum_j b_{ij}^2 + n \sum_i a_i^2 + 2(0) \\ &= SS_{w. \text{people}} + SS_{\text{between people}}. \end{aligned}$$

The difference between two observations on the same person depends in part upon the difference in treatment effects and in part upon uncontrolled or residual sources of variation. Hence the pooled within-person variation may be divided into two parts: one part which depends upon differences between the treatment means, and a second part which consists of residual variation. That part which depends upon differences between treatment effects is defined as

$$(4) \quad SS_{\text{treat}} = n \sum_j (\bar{T}_j - \bar{G})^2.$$

Alternatively, this source of variation may be expressed as

$$SS_{\text{treat}} = \frac{n \sum_j (\bar{T}_j - \bar{T}_{j'})^2}{k(k - 1)}.$$

The expression $\bar{T}_j - \bar{T}_{j'}$ represents the difference between a pair of treatment means; the summation is with respect to all possible pairs of treatment means, order within the pair being disregarded. For example, if $k = 3$,

$$SS_{\text{treat}} = \frac{n[(\bar{T}_1 - \bar{T}_2)^2 + (\bar{T}_1 - \bar{T}_3)^2 + (\bar{T}_2 - \bar{T}_3)^2]}{3(2)}.$$

This source of variation has $k - 1$ degrees of freedom.

The residual variation is

$$(5) \quad SS_{\text{res}} = \sum \sum [(X_{ij} - \bar{G}) - (\bar{P}_i - \bar{G}) - (\bar{T}_j - \bar{G})]^2.$$

The terms that are subtracted from $X_{ij} - \bar{G}$ are, respectively, the person and treatment effects so that the residual variation represents those sources of

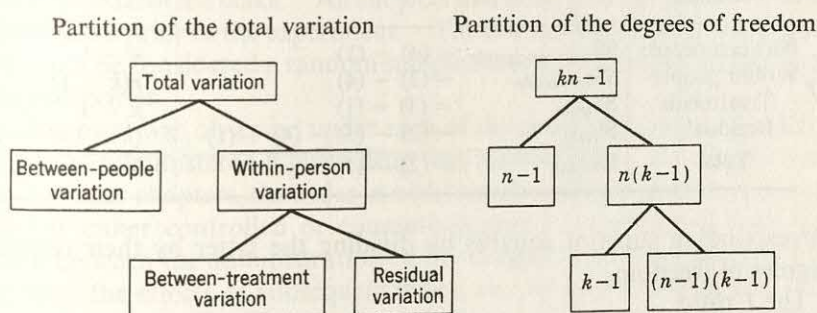


Figure 4.1 Schematic representation of the analysis.

variation in the total that cannot be accounted for by differences between the people and differences between the treatments. The degrees of freedom for the residual variation are

$$\begin{aligned} df_{\text{res}} &= df_{\text{total}} - df_{\text{between people}} - df_{\text{treat}} \\ &= (kn - 1) - (n - 1) - (k - 1) \\ &= kn - n - k + 1 = n(k - 1) - (k - 1) \\ &= (k - 1)(n - 1). \end{aligned}$$

It is readily shown that SS_{treat} and SS_{res} are statistically independent and that

$$SS_{\text{w. people}} = SS_{\text{treat}} + SS_{\text{res}}.$$

The degrees of freedom for the corresponding variations are also additive, i.e.,

$$n(k - 1) - (k - 1) + (n - 1)(k - 1).$$

The analysis of the sources of variation and the corresponding degrees of freedom is shown schematically in Fig. 4.1.

The definitions of the sources of variation do not provide the most convenient formulas for their computation. Formulas for this purpose are summarized in Table 4.2-2. The symbols (1), (2), and (3) are identical to

those used in the case of single-factor experiments which do not have repeated measures. Symbol (4) occurs only in experiments having repeated measures. In each case the divisor in a term is the number of observations that are summed to obtain an element in the numerator. For example, G is the sum of kn observations; T_j is the sum of n observations; P_i is the sum of k observations. A summary of the analysis of variance appropriate for this design is given in part ii of this table. Mean squares are obtained from

Table 4.2-2 Summary of Computational Procedures

(i)		(1) = G^2/kn	(2) = $\Sigma \Sigma X^2$	(3) = $(\Sigma T_j^2)/n$	(4) = $(\Sigma P_i^2)/k$
		Source of variation	SS	df	
(ii)	Between people	$SS_{\text{between people}}$	$= (4) - (1)$	$n - 1$	
	Within people	$SS_{\text{w. people}}$	$= (2) - (4)$	$n(k - 1)$	
	Treatments	SS_{treat}	$= (3) - (1)$	$k - 1$	
	Residual	SS_{res}	$= (2) - (3) - (4) + (1)$	$(n - 1)(k - 1)$	
	Total	SS_{total}	$= (2) - (1)$	$kn - 1$	

corresponding sums of squares by dividing the latter by their respective degrees of freedom.

The F ratio

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}}$$

provides a test of the hypothesis that $\tau_1 = \tau_2 = \dots = \tau_k$, where the τ 's represent treatment effects and are defined in the same manner as they were for the case of designs not having repeated measures. The rationale underlying the use of this statistic for this test is discussed in Sec. 4.4.

Under one set of assumptions (made explicit in Sec. 4.4) about the underlying sources of variation, the F ratio has a sampling distribution which is approximated by the F distribution having $k - 1$ and $(n - 1)(k - 1)$ degrees of freedom. This is the usual test. Under less restrictive assumptions about the relations between the underlying sources of variation, Box (1954) has shown that the F ratio in the last paragraph has a sampling distribution (assuming that all $\tau_j = 0$) which is approximated by the F distribution having $(k - 1)\theta$ and $(n - 1)(k - 1)\theta$ degrees of freedom, where θ is a quantity which depends upon certain homogeneity assumptions. The maximum value of θ is 1.00, and the minimum value is $1/(k - 1)$. The maximum value of θ is attained when the homogeneity assumptions underlying the usual test are met. Use of the minimum value of θ provides a conservative test. Thus, if the ratio

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}}$$

is assumed to be distributed as an F distribution with 1 and $n - 1$ degrees of freedom (assuming that all $\tau_j = 0$), one has a conservative test relative to the usual test. However, the assumptions underlying this test are much weaker than those underlying the usual test. (*Conservative* in this context implies that a larger value of the F ratio is required for statistical significance at a specified level of α .)

4.3 Numerical Example

The computational procedures described in the last section will be illustrated by means of the numerical example in Table 4.3-1. The statistical basis for the analysis is discussed in the next section. The purpose of this experiment was to study the effects of four drugs upon reaction time to a series of standardized tasks. All subjects had been given extensive training on these tasks prior to the experiment. The five subjects used in the experiment could be considered a random sample from a population of interest to the experimenter.

Each subject was observed under each of the drugs; the order in which a subject was administered a given drug was randomized. (In designs considered in later chapters, the order in which treatments are given to the same subject is either controlled or counterbalanced.) A sufficient time was allowed between the administration of the drugs to avoid the effect of one drug upon the effects of subsequent drugs, i.e., an interaction effect. The numerical entries in Table 4.3-1 represent the score (mean reaction time) on the series of standardized tasks. Thus person 1 had scores of 30, 28, 16, and 34 under the respective drug conditions. The total of these scores is 108; thus the numerical value of P_1 is 108. The other values for the P 's are obtained by summing the entries in the respective rows in part i. The numerical values for the T 's are obtained by summing the columns. For example, T_1 is the sum of the five entries under drug 1. The grand total, G , is obtained either by summing the P 's or by summing the T 's. A check on the arithmetic work is provided by computing G by both methods.

Quantities required in the computation of the sums of squares are given in part ii. The first three of these quantities are identical to those computed for designs which do not involve repeated measures. Symbol (4) is obtained from the P 's. Each P is the sum over $k = 4$ drugs; hence the divisor associated with the symbol (4) is 4. The computation of the sums of squares required in the analysis of variance is illustrated in part iii. An alternative method for computing SS_{res} is

$$\begin{aligned} SS_{\text{res}} &= SS_{\text{w. people}} - SS_{\text{drugs}} \\ &= 811.00 - 698.20 = 112.80. \end{aligned}$$

The latter method is actually simpler than the method used in part iii; however, the method in part iii provides a partial check on the numerical work, since the sum of SS_{drugs} and SS_{res} should total $SS_{\text{w. people}}$.

The analysis of variance is summarized in Table 4.3-2. The F ratio

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}} = \frac{232.73}{9.40} = 24.76$$

is used in testing hypotheses about reaction time as a function of the effects of the drugs. For a .01-level test on the hypothesis that $\tau_1 = \tau_2 = \tau_3 = \tau_4$, the critical value for the F ratio is $F_{.99}(3,12) = 5.95$. The experimental

Table 4.3-1 Numerical Example

Person	Drug 1	Drug 2	Drug 3	Drug 4	Total
1	30	28	16	34	108 = P_1
2	14	18	10	22	64 = P_2
(i) 3	24	20	18	30	92 = P_3
4	38	34	20	44	136 = P_4
5	26	28	14	30	98 = P_5
	132 T_1	128 T_2	78 T_3	160 T_4	498 = G

$$(1) = \frac{G^2}{kn} = \frac{(498)^2}{4(5)} = \frac{248,004}{20} = 12,400.20$$

$$(2) = \Sigma \Sigma X^2 = 13,892$$

$$(ii) (3) = \frac{\Sigma T_j^2}{n} = \frac{132^2 + 128^2 + 78^2 + 160^2}{5} = \frac{65,492}{5} = 13,098.40$$

$$(4) = \frac{\Sigma P_i^2}{k} = \frac{108^2 + 64^2 + 92^2 + 136^2 + 98^2}{4} = \frac{52,324}{4} = 13,081.00$$

$$SS_{\text{between people}} = (4) - (1) = 13,081.00 - 12,400.20 = 680.80$$

$$SS_{\text{w. people}} = (2) - (4) = 13,892 - 13,081.00 = 811.00$$

$$(iii) SS_{\text{drugs}} = (3) - (1) = 13,098.40 - 12,400.20 = 698.20$$

$$SS_{\text{res}} = (2) - (3) - (4) + (1) = 13,892 - 13,098.40 - 13,081.00 + 12,400.20 = 112.80$$

$$SS_{\text{total}} = (2) - (1) = 13,892 - 12,400.20 = 1491.80$$

data contradict this hypothesis. Inspection of the totals for the drugs in Table 4.3-1 indicates that drug 3 is associated with the fastest reaction.

Suppose that it had been anticipated before the experiment had been conducted that drug 3 would have a different effect from all others. The comparison that would be used in testing this hypothesis is

$$C = 3T_3 - T_1 - T_2 - T_4 = 3(78) - 132 - 128 - 160 = -186.$$

The component of variation corresponding to this comparison is

$$SS_C = \frac{C^2}{n\Sigma c^2} = \frac{(-186)^2}{5[3^2 + (-1)^2 + (-1)^2 + (-1)^2]} = 576.60.$$

The F statistic

$$F = \frac{SS_C}{MS_{\text{res}}} = \frac{576.60}{9.40} = 61.34$$

is used to test the hypothesis that $\tau_3 = (\tau_1 + \tau_2 + \tau_4)/3$. The critical value for a .01-level test of this hypothesis $F_{.99}(1,12) = 9.33$. The observed data contradict this hypothesis. If this comparison were suggested by inspection of the data, the procedure given by Scheffé (described in Sec. 3.9) would be used to obtain the critical value. The latter critical value for a .01-level test is $(k-1)F_{.99}(k-1, df_{\text{res}}) = 3F_{.99}(3,12) = 3(5.95) = 17.85$. Even

Table 4.3-2 Analysis of Variance

Source of variation	SS	df	MS	F
Between people	680.80	4		
Within people	811.00	15		
Drugs	698.20	3	232.73	24.76**
Residual	112.80	12	9.40	
Total	1491.80	19		

** $F_{.99}(3,12) = 5.95$

with this critical value, the data indicate that drug 3 is different in its effect on reaction time from the effects of the other three drugs.

To test the hypothesis that $\tau_1 = \tau_2 = \tau_4$, the sum of squares for these three drugs is given by

$$\begin{aligned} SS_{\text{drugs 1, 2, 4}} &= \frac{T_1^2 + T_2^2 + T_4^2}{n} - \frac{(T_1 + T_2 + T_4)^2}{3n} \\ &= \frac{132^2 + 128^2 + 160^2}{5} - \frac{(132 + 128 + 160)^2}{15} \\ &= 121.60. \end{aligned}$$

The mean square corresponding to this sum of squares is

$$MS_{\text{drugs 1, 2, 4}} = \frac{121.60}{2} = 60.80.$$

The statistic used in the test is

$$F = \frac{MS_{\text{drugs 1, 2, 4}}}{MS_{\text{res}}} = \frac{60.80}{9.40} = 6.47.$$

For a .01-level test, the critical value of this statistic is $F_{.99}(2,12) = 6.93$. Although the observed F statistic does not exceed the critical value for a .01-level test, the observed F is large enough to question the hypothesis that the drugs 1, 2, and 4 are equally effective with respect to reaction time. Inspection of the drug totals in Table 4.3-1 indicates that drug 4 has a somewhat longer reaction time, but the evidence is not quite strong enough to establish this conclusion at the .01-level of significance.

The data, in this case, can be adequately summarized in terms of a few selected comparisons. Analogous conclusions can be reached by other, somewhat more systematic probing procedures. Any of the methods discussed in Secs. 3.8 and 3.9 may be used to test the difference between all possible pairs of means. In such tests MS_{res} has the role of MS_{error} . Application of the Newman-Keuls method is illustrated in Table 4.3-3.

Table 4.3-3 Tests on Differences between Pairs of Means

	Drugs				
		3	2	1	4
(i)	Totals	78	128	132	160
	3	78	—	50	54
	2	128	—	4	32
	1	132	—	—	28
	4	160	—	—	—
(ii)	$q_{.99}(r, 12)$	4.32	5.04	5.50	
(iii)	$\sqrt{nMS_{\text{res}}} q_{.99}(r, 12)$	29.64	34.57	37.73	
(iv)		3	2	1	4
	3		**	**	**
	2		—	—	—
	1			—	—
	4				—

With repeated measures, barring missing data, the numbers of observations under each treatment will be equal. In this case treatment totals may be used rather than treatment means. The drug totals, in increasing order of magnitude, are given in part i. The entry in a cell of part i is the difference between a total at the head of a column and a total to the left of a row. Critical values for the statistic

$$q_r = \frac{T_j - T_{j'}}{\sqrt{nMS_{\text{res}}}},$$

where r is the number of steps two totals are apart on an ordered scale, are given in part ii. These values are obtained from the first three columns of tables for the 99th percentile point for the q statistic; the degrees of freedom for this q statistic are the degrees of freedom of MS_{res} . Critical values for

$$T_j - T_{j'} = q_r \sqrt{nMS_{\text{res}}}$$

are given in part iii. In this case

$$\sqrt{nMS_{\text{res}}} = \sqrt{5(9.40)} = \sqrt{47.00} = 6.86.$$

Thus the entries in part iii are 6.86 times the corresponding entries in part ii.

The order in which tests are made is given in Sec. 3.8. The critical value for the difference $T_4 - T_3 = 82$ is 37.73. Hence the data contradict the hypothesis that $\tau_4 = \tau_3$. The difference $T_1 - T_3 = 54$ has the critical value 34.57, and the difference $T_2 - T_3 = 50$ has the critical value 29.64. The difference $T_4 - T_2 = 32$ has the critical value 34.57; this difference does not quite exceed the critical value. No further tests are made. The tests which yield statistically significant results are summarized in part iv. Drug 3 appears to be different from the other drugs in its effect on reaction time. Although the differences between drug 4 and drugs 2 and 1 are relatively large, the differences do not exceed critical values of a .01-level test. This latter result is consistent with the outcome of the test of the hypothesis that

Table 4.3-4 Variance-Covariance Matrix

	Drug 1	Drug 2	Drug 3	Drug 4
Drug 1	76.80	53.20	29.20	69.00
Drug 2		42.80	15.80	47.00
Drug 3			14.80	27.00
Drug 4				64.00

$\tau_1 = \tau_2 = \tau_4$. This hypothesis was not rejected at the .01 level of significance, but the observed F statistic was close to the critical value.

The computational formula for MS_{res} is algebraically equivalent to the expression

$$MS_{res} = s_X^2 - \overline{cov}_X,$$

where s_X^2 is the mean of the variances within each of the drug conditions and \overline{cov}_X is the mean of the covariances between the pairs of observations under any two drug conditions. To show this equivalence for the numerical data in Table 4.3-1, the variance-covariance matrix for these data is given in Table 4.3-4.

The variance of the observations made under each of the drugs appears along the main diagonal of this table. For example, the variance of the observations made under drug 1 is

$$s_{X_1}^2 = \frac{\Sigma X_1^2 - (T_1^2/n)}{n - 1} = \frac{3792 - (132^2/5)}{4} = 76.80.$$

The covariances appear above the main diagonal. For example, the covariance between the observations under drug 1 and those made under drug 2 is

$$\begin{aligned} cov_{X_1 X_2} &= \frac{\Sigma(X_{i1}X_{i2}) - (T_1T_2/n)}{n - 1} \\ &= \frac{(30)(28) + \cdots + (26)(28) - [(132)(128)/5]}{4} = 53.20. \end{aligned}$$

The mean of the variances is 49.60; the mean of the covariances is 40.20. Thus,

$$\bar{s}_X^2 - \overline{\text{cov}}_X = 49.60 - 40.20 = 9.40.$$

The numerical value of MS_{res} obtained by the computational formula is also 9.40. It is considerably more work to obtain MS_{res} from the variance-covariance matrix than it is to obtain MS_{res} by means of the computational formula. However, in order to check certain of the assumptions underlying the F test, computation of the variance-covariance matrix is sometimes required and is often enlightening in its own right.

$\text{MS}_{\text{between people}}$ is also related to \bar{s}_X^2 and $\overline{\text{cov}}_X$. This relationship is

$$\text{MS}_{\text{between people}} = \bar{s}_X^2 + (k - 1) \overline{\text{cov}}_X.$$

From Table 4.3-2,

$$\text{MS}_{\text{between people}} = \frac{\text{SS}_{\text{between people}}}{k - 1} = \frac{680.80}{4} = 170.20.$$

Table 4.3-5 Interrelation Matrix

	Drug 1	Drug 2	Drug 3	Drug 4
Drug 1	1.000	.917	.860	.978
Drug 2		1.000	.620	.887
Drug 3			1.000	.877
Drug 4				1.000

In terms of the average variance and the average covariance,

$$\text{MS}_{\text{between people}} = 49.60 + 3(40.20) = 170.20.$$

The matrix of intercorrelations for the data in Table 4.3-1 is given in Table 4.3-5. The average intercorrelation of the off-diagonal elements is $\bar{r} = .86$. Had a common estimate of the population variance been used in computing these intercorrelations, this average would have been $\bar{r}' = .81$. In terms of this latter average,

$$\text{MS}_{\text{res}} = \bar{s}_X^2(1 - \bar{r}') = 49.60(1 - .81) = 9.42.$$

Had the data been uncorrelated, MS_{res} would have been equal to 49.60. The larger the average intercorrelation, the smaller will be MS_{res} .

4.4 Statistical Basis for the Analysis

The validity of tests in the last section rests upon a set of assumptions about the nature of the underlying sources of variation. The case in which the number of treatments is equal to 2 will be considered first. The assumptions that will be made for this case are actually stronger than those required; however, all the assumptions will be required for the case in which the

number of treatments is greater than 2. Suppose that each of the observations may be expressed in terms of the strictly additive model given below:

Person	Treatment 1	Treatment 2
1	$X_{11} = \mu + \pi_1 + \tau_1 + \varepsilon_{11}$	$X_{12} = \mu + \pi_1 + \tau_2 + \varepsilon_{12}$
.	.	.
.	.	.
n	$X_{n1} = \mu + \pi_n + \tau_1 + \varepsilon_{n1}$	$X_{n2} = \mu + \pi_n + \tau_2 + \varepsilon_{n2}$
	$T_1 = n\mu + \sum \pi_i + n\tau_1 + \sum \varepsilon_{i1}$ $\bar{T}_1 = \mu + \bar{\pi} + \tau_1 + \bar{\varepsilon}_2$	$T_2 = n\mu + \sum \pi_i + n\tau_2 + \sum \varepsilon_{i2}$ $\bar{T}_2 = \mu + \bar{\pi} + \tau_2 + \bar{\varepsilon}_2$

The notation is defined below:

X_{ij} = an observation on person i under treatment j .

μ_1 = mean of all potential observations under treatment 1, that is, mean for treatment 1 if the entire population of people were observed under treatment 1.

μ_2 = mean of all potential observations under treatment 2.

π_i = a constant associated with person i . In the population of people the mean of the π_i is assumed to be zero.

$\mu = (\mu_1 + \mu_2)/2$ = grand mean of all potential observations.

$\tau_1 = \mu_1 - \mu$ = main effect of treatment 1.

$\tau_2 = \mu_2 - \mu$ = main effect of treatment 2.

ε_{ij} = experimental error associated with X_{ij} .

= all sources of variation in X_{ij} except those accounted for by the τ 's and the π 's.

Treatments 1 and 2 will be assumed to constitute the population of treatments. From the definition of τ_1 and τ_2 , it follows that $\tau_1 + \tau_2 = 0$. Since the n people in the experiment are assumed to be a random sample from a potentially infinite population of people, π_i is a random variable. In the population of people, π_i will be assumed to be normally distributed with mean zero and variance σ_π^2 .

Within the population of potential observations under treatment 1, the experimental error ε_{i1} is assumed to be normally distributed, with mean equal to zero and variance equal to $\sigma_{\varepsilon_1}^2$. Within the corresponding population under treatment 2, the experimental error is also assumed to be normally distributed, with mean equal to zero and variance equal to $\sigma_{\varepsilon_2}^2$. The two distributions of experimental error will be assumed to be independent; further it will be assumed that

$$\sigma_{\varepsilon_1}^2 = \sigma_{\varepsilon_2}^2 = \sigma_\varepsilon^2;$$

i.e., the error variances are homogeneous.

Since μ and τ_1 are constant for all observations under treatment 1, the variance of the X 's within treatment population 1 is a function of the variance due to π_i and ε_{i1} . Assuming π_i and ε_{i1} uncorrelated,

$$\sigma_{X_1}^2 = \sigma_\varepsilon^2 + \sigma_\pi^2.$$

In words, under the assumptions that have been made, the variance of the potential observations under treatment 1 is the sum of the variance due to experimental error and the variance due to differences between the π 's. Similarly the variance due to the potential observations under treatment 2 is

$$\sigma_{X_2}^2 = \sigma_\varepsilon^2 + \sigma_\pi^2.$$

Since the term π_i is common to two measurements on the same person, the covariance between X_1 and X_2 will not in general be equal to zero. All the covariance between X_1 and X_2 is assumed to be due to the term π_i . If this covariance is denoted by the symbol $\sigma_{X_1X_2}$, then

$$\sigma_{X_1X_2} = \sigma_\pi^2.$$

Under the assumptions that have been made,

$$\frac{\sigma_{X_1}^2 + \sigma_{X_2}^2}{2} = \sigma_\varepsilon^2 + \sigma_\pi^2.$$

Hence

$$\frac{\sigma_{X_1}^2 + \sigma_{X_2}^2}{2} - \sigma_{X_1X_2} = \sigma_\varepsilon^2.$$

The computational formula for MS_{res} , given in Sec. 4.2, is equivalent to the left-hand side of the above expression if statistics are substituted for corresponding parameters. MS_{res} provides an unbiased estimate of σ_ε^2 ; in symbols, this last statement is expressed by

$$E(MS_{res}) = \sigma_\varepsilon^2.$$

The expectation in this case is with respect to random samples of size n . Under the assumptions made, the sampling distribution of $(n-1)MS_{res}/\sigma_\varepsilon^2$ may be approximated by a chi-square distribution having $n-1$ degrees of freedom.

From the structural model it is seen that

$$\bar{T}_1 - \bar{T}_2 = (\tau_1 - \tau_2) + (\bar{\varepsilon}_1 - \bar{\varepsilon}_2).$$

Since the same people are observed under both the treatments, this difference is free of any effects associated with the π 's. The variance of the quantity $\bar{T}_1 - \bar{T}_2$, when the experiment is replicated with random samples of size n people, has the form

$$\sigma_{\bar{T}_1 - \bar{T}_2}^2 = \sigma_{\tau_1 - \tau_2}^2 + \sigma_{\bar{\varepsilon}_1 - \bar{\varepsilon}_2}^2.$$

This expression assumes that $\tau_1 - \tau_2$ and $\bar{\varepsilon}_1 - \bar{\varepsilon}_2$ are uncorrelated. An equivalent expression for $\sigma_{\bar{T}_1 - \bar{T}_2}^2$ is

$$\sigma_{\bar{T}}^2 = \sigma_\tau^2 + \sigma_{\bar{\varepsilon}}^2.$$

This last expression is implied by the previous expression, since the variance of a variable is a function of the differences between all possible pairs of the variables. Multiplying both sides of this last expression by n gives

$$n\sigma_T^2 = n\sigma_r^2 + n\sigma_e^2.$$

For random samples of size n , it has been shown in earlier chapters that $n\sigma_e^2 = \sigma_e^2$. It may further be shown that MS_{treat} is an unbiased estimate of $n\sigma_T^2$. Hence

$$E(MS_{\text{treat}}) = n\sigma_T^2 = n\sigma_r^2 + \sigma_e^2.$$

Under the assumptions that have been made, when $\sigma_r^2 = 0$ the sampling distribution of $MS_{\text{treat}}/\sigma_e^2$ may be approximated by a chi-square distribution having one degree of freedom. Further the sampling distribution of MS_{treat} is independent of the sampling distribution of MS_{res} . Thus the statistic

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}},$$

under the hypothesis that $\sigma_r^2 = 0$, represents the ratio of mean squares having independent chi-square distributions. Further, when $\sigma_r^2 = 0$, the numerator and denominator have expected values equal to σ_e^2 . Hence the F statistic has a sampling distribution which may be approximated by an F distribution having one degree of freedom for the numerator and $n - 1$ degrees of freedom for the denominator. More explicitly,

$$E(F) = \frac{\sigma_e^2 + n\sigma_r^2}{\sigma_e^2} = \frac{\sigma_e^2}{\sigma_e^2} = 1.00,$$

when $\sigma_r^2 = 0$.

The model under which this last F ratio was obtained did not include a possible person by treatment interaction. The more inclusive model is

$$X_{ij} = \mu + \pi_i + \tau_j + \pi\tau_{ij} + \varepsilon_{ij}.$$

The person by treatment interaction, $\pi\tau_{ij}$, is a measure of the reliable effect associated with person i which is exhibited in the presence of treatment j and which cannot be predicted from knowledge of π_i and τ_j . The interaction effect $\pi\tau_{ij}$ is considered to be a random variable, normally distributed, with mean equal to zero within each of the treatment populations, and with variance

$$\sigma_{\pi\tau_1}^2 = \sigma_{\pi\tau_2}^2 = \sigma_{\pi\tau}^2.$$

If the $\pi\tau_{ij}$'s are defined to be uncorrelated (a less restrictive assumption is made later in this section) with the other terms on the right-hand side of the model, it may be shown that

$$E(MS_{\text{treat}}) = \sigma_e^2 + \sigma_{\pi\tau}^2 + n\sigma_r^2,$$

$$E(MS_{\text{res}}) = \sigma_e^2 + \sigma_{\pi\tau}^2.$$

The sampling distribution of the ratio of these two mean squares is still approximately an F distribution with one and $n - 1$ degrees of freedom.

In this design σ_e^2 and $\sigma_{\pi\tau}^2$ cannot be estimated separately. Only their sum may be estimated. Since they cannot be estimated separately, there is no real need to have both effects in the model; if there is a person by treatment effect, it is completely confounded with the residual or experimental error. In designs which are discussed in later chapters, it is possible to obtain a direct estimate of σ_e^2 .

There are alternative approaches which justify the use of the final F ratio as a test of the hypothesis that $\sigma_\tau^2 = 0$. However, the approach that has been followed leads more directly to the general case in which the number of treatments is any number k . For this general case, an observation X_{ij} on person i under treatment j may be expressed in terms of the strictly additive model as

$$X_{ij} = \mu + \pi_i + \tau_j + \varepsilon_{ij}.$$

The terms on the right-hand side of this model have definitions analogous to those given for the case in which $k = 2$. The variables π_i and ε_{ij} are assumed to be normally distributed, with expected value equal to zero within each of the treatment populations. It is also assumed that π_i and ε_{ij} are uncorrelated. Since the same people are observed under each of the treatment populations, σ_π^2 is assumed to be the same for all treatments. It will also be assumed that the variance due to experimental error, σ_e^2 , is the same for all treatment populations.

The subjects in a single experiment are assumed to be a random sample of size n from a population of N people, where N is quite large relative to n . The population variance for treatment j is defined as (the summation is over the population of subjects)

$$\sigma_{X_j}^2 = \frac{\sum_i (X_{ij} - \mu_j)^2}{N - 1},$$

where $\mu_j = \sum_i X_{ij}/N$ is the population mean for treatment j . The population covariance between X_{ij} and $X_{ij'}$, where j and j' represent two different treatments, is defined as

$$\begin{aligned}\sigma_{X_j X_{j'}} &= \frac{\sum_i (X_{ij} - \mu_j)(X_{ij'} - \mu_{j'})}{N - 1} \\ &= \rho_{X_j X_{j'}} \sigma_{X_j} \sigma_{X_{j'}}.\end{aligned}$$

The symbol $\rho_{X_j X_{j'}}$ denotes the correlation between observations in treatment population j and treatment population j' . (In designs which do not involve repeated measures on the same people this correlation will be zero.)

If the underlying model is correct, the covariance between any two treatments is due solely to the presence of π_i in pairs of measures on the same people. From this it follows that all the population covariances will be equal; i.e.,

$$\sigma_{X_i X_{j'}} = \sigma^2 \quad \text{for all treatment populations.}$$

Further, since π_i and ε_{ij} have been assumed to be uncorrelated, with σ_ε^2 being constant for all treatment populations, it follows that the variance of the potential observations within each of the treatment populations may be expressed as

$$\sigma_X^2 = \sigma_\varepsilon^2 + \sigma_\pi^2.$$

Hence

$$\sigma_\varepsilon^2 = \sigma_X^2 - \sigma_\pi^2 = \sigma_X^2 - \sigma_{X_j X_{j'}}.$$

In words, the variance attributable to experimental error is equal to the within-treatment variance minus the covariance between treatments.

Each of the terms in this last expression is a parameter. Unbiased estimates of these parameters are available from experimental data, provided that experimental data are obtained under conditions which approximate those assumed by the structural model. The mean of the within-treatment variances for the experimental data provides an unbiased estimate of σ_X^2 ; that is,

$$E(\bar{s}_X^2) = E\left(\frac{s_1^2 + s_2^2 + \cdots + s_k^2}{k}\right) = \sigma_X^2.$$

Similarly the mean of all the covariances will provide an unbiased estimate of $\sigma_{X_j X_{j'}}$. The mean of these covariances will be designated by the symbol $\overline{\text{cov}}$. It may further be shown that

$$E(\bar{s}_X^2 - \overline{\text{cov}}) = \sigma_\varepsilon^2.$$

The computational formula for MS_{res} given in Sec. 4.2 may be shown to be algebraically equivalent to the expression in parentheses. Thus

$$\text{MS}_{\text{res}} = \bar{s}_X^2 - \overline{\text{cov}}.$$

Hence

$$E(\text{MS}_{\text{res}}) = \sigma_\varepsilon^2.$$

The expectation in this case is with respect to replications of the experiment with random samples of n subjects. Under the conditions that have been specified, the sampling distribution of $(k-1)(n-1)\text{MS}_{\text{res}}/\sigma_\varepsilon^2$ may be approximated by a chi-square distribution having $(k-1)(n-1)$ degrees of freedom.

The difference between the mean of the n observations under treatment j , \bar{T}_j , and the mean of the n observations under treatment j' , $\bar{T}_{j'}$, may be expressed in terms of the linear model as

$$\bar{T}_j - \bar{T}_{j'} = (\tau_j - \tau_{j'}) + (\bar{\varepsilon}_j - \bar{\varepsilon}_{j'}).$$

Since the same people appear under both the treatments, a term involving the difference between two π 's does not appear in the above expression. Since it has been assumed that the treatment effects and the experimental error are uncorrelated, the variance of the difference on the left-hand side of the above expression will provide an estimate of the sum of the variances of the differences on the right-hand side. In symbols,

$$\text{MS}_{\text{treat}} = \frac{n \sum (\bar{T}_j - \bar{T}_{j'})^2}{k(k-1)} = \sigma_\varepsilon^2 + n\sigma_\tau^2.$$

The computational formula for MS_{treat} given in Sec. 4.2 is algebraically equivalent to that appearing on the left-hand side of the above expression.

Under the hypothesis that $\sigma_\tau^2 = 0$, the sampling distribution of $(k-1)MS_{\text{treat}}/\sigma_\epsilon^2$ may be approximated by a chi-square distribution having $k-1$ degrees of freedom. Further, under this hypothesis both MS_{treat} and MS_{res} provide independent estimates of σ_ϵ^2 . Hence the F ratio

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}}$$

has a sampling distribution approximated by an F distribution having $k-1$ and $(k-1)(n-1)$ degrees of freedom. If σ_τ^2 were greater than zero, the numerator of the F statistic would have an expected value which differs from the expected value of the denominator by an amount equal to $n\sigma_\tau^2$. Hence numerically large values of the F ratio will tend to contradict the hypothesis that $\sigma_\tau^2 = 0$. In terms of expected values,

$$E(F) = \frac{E(MS_{\text{treat}})}{E(MS_{\text{res}})} = \frac{\sigma_\epsilon^2 + n\sigma_\tau^2}{\sigma_\epsilon^2} = 1.00 \quad \text{when } \sigma_\tau^2 = 0.$$

In the structural model from which the expected values were obtained, no person \times treatment interaction appears. If this interaction is defined as

$$\pi\tau_{ij} = E(X_{ij}) - \tau_j - \pi_i,$$

then the structural model has the form

$$X_{ij} = \mu + \tau_j + \pi_i + \pi\tau_{ij} + \epsilon_{ij}.$$

In this model ϵ_{ij} is assumed to be normally distributed with variance σ_ϵ^2 within each of the treatment populations and statistically independent of the other terms on the right-hand side; τ_j is assumed to be a constant associated with treatment j ; π_i and $\pi\tau_{ij}$ are assumed to be random variables having a joint normal distribution.

In order that the usual F test on the main effects due to the treatments be valid, the following restrictions are imposed upon the model:

$$\sigma_{X_j}^2 = \sigma_{X_{j'}}^2 = \sigma_X^2,$$

$$\sigma_{X_j X_{j'}} = \rho \sigma_X^2 \quad \text{for all } j\text{'s and } j'\text{'s.}$$

Under these conditions it can be shown that

$$E(MS_{\text{between people}}) = \sigma_\epsilon^2 + k\sigma_\pi^2 = \sigma_X^2[1 + (k-1)\rho],$$

$$E(MS_{\text{treat}}) = \sigma_\epsilon^2 + \sigma_{\pi\tau}^2 + n\sigma_\tau^2 = \sigma_X^2(1 - \rho) + n\sigma_\tau^2,$$

$$E(MS_{\text{res}}) = \sigma_\epsilon^2 + \sigma_{\pi\tau}^2 = \sigma_X^2(1 - \rho).$$

Hence

$$MS_{\text{between people}} = \bar{s}_X^2 + (k-1)\overline{\text{cov}},$$

$$MS_{\text{res}} = \bar{s}_X^2 - \overline{\text{cov}}.$$

A formal proof of the equivalence of these expected values is given by

Scheffé (1960, pp. 262–267). The F ratio in the test of the hypothesis $\sigma^2_{\tau} = 0$ still has the form

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}}.$$

The assumptions of homogeneity of within-treatment population variance $\sigma^2_{X_1} = \sigma^2_{X_2} = \cdots = \sigma^2_{X_k}$ and homogeneity of covariance

$$\sigma_{X_1 X_2} = \sigma_{X_1 X_3} = \cdots = \sigma_{X_{(k-1)} X_k},$$

which underlie the validity of the usual F test for this design, are highly restrictive. That is, the data must conform to a prescribed pattern of variances and covariances before the statistical test can be considered exact. Box (1953) has indicated that the usual F test in case of uncorrelated data is relatively robust (insensitive) with respect to violation of the assumption of homogeneity of variance. That is, for uncorrelated data, violation of the assumption of homogeneity of variance does not seriously bias the final F test.

Box has shown, however, that heterogeneity of both the variances and covariances in a design having correlated observations will generally result in a positive bias in the usual F test. That is, the critical value as obtained from an F table tends to be too low relative to a critical value appropriate for an arbitrary variance-covariance matrix. (The usual F test is appropriate only for a very special variance-covariance matrix.) For the case in which the variance-covariance matrix is arbitrary in form, an approximate test may be made through use of the usual F statistic, but the degrees of freedom are taken to be $(k-1)\theta$ and $(k-1)(n-1)\theta$, where θ is a number which depends upon the degree of heterogeneity of the variances and covariances. When the assumptions of homogeneity are met, θ equals unity. The greater the heterogeneity, the smaller θ is. The smallest possible value that θ can assume is $1/(k-1)$.

If one assumes that θ is actually equal to this lower bound, the resulting test will tend to err on the conservative side. That is, the F value determined from an F table will tend to be somewhat larger than the exact value. (This kind of test is said to be negatively biased.) The degrees of freedom for the numerator of the F ratio, assuming that $\theta = 1/(k-1)$, are

$$(k-1)\theta = \frac{k-1}{k-1} = 1.$$

The degrees of freedom for the denominator, assuming that $\theta = 1/(k-1)$, are

$$(n-1)(k-1)\theta = \frac{(n-1)(k-1)}{k-1} = n-1$$

Hence the conservative test assumes that the F ratio has one and $n-1$ degrees of freedom. When the assumptions of homogeneity of variances and homogeneity of covariances are questionable, the conservative test indicated in this paragraph provides an approximate test.

For the case in which $k > 2$ and $n > k$, Hotelling's T^2 statistic may be used to test the hypothesis that $\sigma_r^2 = 0$. This test is exact if the underlying distributions are multivariate normal. Use of Hotelling's T^2 statistic requires no assumptions of homogeneity of covariances on the population variance-covariance matrix. Computationally, Hotelling's T^2 statistic is more complex, since it involves the inverse of the covariance matrix. When the homogeneity assumptions are met, the T^2 statistic is algebraically equivalent to the usual F statistic. For a more detailed discussion of this statistic see Anderson (1958, pp. 101-105) and Geisser (1959). A numerical example of the T^2 statistic appears in Appendix A.6.

Procedures for testing homogeneity hypotheses about population covariance matrices are described in Sec. 7.7. In particular χ^2_2 as defined in Sec. 7.7 may be used to test the hypothesis that the population covariance matrix has the form

$$\begin{bmatrix} \sigma^2 & \rho\sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \rho\sigma^2 \\ \rho\sigma^2 & \rho\sigma^2 & \sigma^2 \end{bmatrix}.$$

This is the form specified by the model used in this section to justify the subsequent F tests. In the definition of χ^2_2 in Sec. 7.7, q corresponds to the number of treatments; in the notation of this section, $k = q$, and $N = nk$.

4.5 Use of Analysis of Variance to Estimate Reliability of Measurements

Given a person possessing a magnitude π of a specified characteristic. In appraising this characteristic with some measuring device, the observed score may have the magnitude $\pi + \eta$. The quantity η is the error of measurement; all measurement has some of this kind of error. The latter is due in part to the measuring device itself and in part to the conditions surrounding the measurement. In the development that follows, it will be assumed that the magnitude of the error of measurement is uncorrelated with π . A measurement on person i with measuring instrument j may be represented as

$$(1) \quad X_{ij} = \pi_i + \eta_{ij},$$

where X_{ij} = observed measurement,

π_i = true magnitude of characteristic being measured,

η_{ij} = error of measurement.

Upon repeated measurement with the same or comparable instruments, π_i is assumed to remain constant, whereas η_{ij} is assumed to vary. The mean of k such repeated measures may be represented as

$$(2) \quad \frac{\sum_j X_{ij}}{k} = \bar{P}_i = \pi_i + \bar{\eta}_i.$$

A schematic representation of a random sample of k measurements on the same or comparable measuring instruments is shown in Table 4.5-1. If π_i remains constant for such measurement, the variance within person i is due to error of measurement and the pooled within-person variance also estimates variance due to error of measurement. On the other hand, the variance in the \bar{P} 's is in part due to differences between the true magnitudes of the characteristic possessed by the n people and in part due to differences in the average error of measurement for each person. The analysis of variance

Table 4.5-1 Estimation of Reliability

Person	Comparable measurements						Total	Mean
	1	2	...	j	...	k		
1	X_{11}	X_{12}		X_{1j}		X_{1k}	P_1	\bar{P}_1
2	X_{21}	X_{22}		X_{2j}		X_{2k}	P_2	\bar{P}_2
.
.
.
i	X_{i1}	X_{i2}		X_{ij}		X_{ik}	P_i	\bar{P}_i
.
.
.
n	X_{n1}	X_{n2}		X_{nj}		X_{nk}	P_n	\bar{P}_n
Total	T_1	T_2	...	T_j	...	T_k	G	

and the expected values for the mean squares for data of the type shown in Table 4.5-1 are given in Table 4.5-2. $MS_{\text{between people}}$ is defined as

$$MS_{\text{between people}} = \frac{k \sum (\bar{P}_i - \bar{G})^2}{n - 1},$$

whereas the variance of the \bar{P} 's is given by

$$s_{\bar{P}}^2 = \frac{\sum (\bar{P}_i - \bar{G})^2}{n - 1}.$$

Thus

$$MS_{\text{between people}} = k s_{\bar{P}}^2.$$

In terms of (2), the expected value of the variance of the \bar{P} 's is

$$E(s_{\bar{P}}^2) = \sigma_{\bar{\eta}}^2 + \sigma_{\pi}^2.$$

The quantity σ_{π}^2 is the variance of the true measures in the population of which the n people in the study represent a random sample. From the relationship between $MS_{\text{between people}}$ and $s_{\bar{P}}^2$,

$$E(MS_{\text{between people}}) = k \sigma_{\bar{\eta}}^2 + k \sigma_{\pi}^2 = \sigma_{\eta}^2 + k \sigma_{\pi}^2,$$

since

$$k \sigma_{\bar{\eta}}^2 = \sigma_{\eta}^2.$$

The reliability of \bar{P}_i , which is the mean of k measurements, is defined to be

$$(3) \quad \rho_k = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\eta^2} = \frac{\sigma_\pi^2}{\sigma_\pi^2 + (\sigma_\eta^2/k)}.$$

In words, the reliability of the mean of k measurements is the variance due to the true scores divided by the sum of the variance due to the true scores and variance due to error of measurement. From Table 4.5-2, it is noted that $MS_{w. \text{ people}}$ provides an estimate of σ_η^2 and $MS_{\text{between people}}$ provides an estimate of $\sigma_\eta^2 + k\sigma_\pi^2$. Hence

$(MS_{\text{between people}} - MS_{w. \text{ people}})$ estimates $k\sigma_\pi^2$,

and $\frac{1}{k} (MS_{\text{between people}} - MS_{w. \text{ people}})$ estimates σ_π^2 .

Table 4.5-2 Analysis of Variance

Source of variation	MS	E(MS)
Between people	$MS_{\text{between people}}$	$\sigma_\eta^2 + k\sigma_\pi^2$
Within people	$MS_{w. \text{ people}}$	σ_η^2

Thus, the estimated reliability for the mean of k measurements, upon substituting in (3), is

$$(4) \quad r_k = \frac{(1/k)(MS_{\text{between people}} - MS_{w. \text{ people}})}{(1/k)(MS_{\text{between people}} - MS_{w. \text{ people}}) + (1/k)MS_{w. \text{ people}}} \\ = \frac{MS_{\text{between people}} - MS_{w. \text{ people}}}{MS_{\text{between people}}} \\ = 1 - \frac{MS_{w. \text{ people}}}{MS_{\text{between people}}}.$$

The reliability of X_{ij} , which is a single measurement, is defined as

$$(5) \quad \rho_1 = \frac{\sigma_\pi^2}{\sigma_\pi^2 + \sigma_\eta^2}.$$

The estimate of this reliability is

$$(6) \quad r_1 = \frac{(1/k)(MS_{\text{between people}} - MS_{w. \text{ people}})}{(1/k)(MS_{\text{between people}} - MS_{w. \text{ people}}) + MS_{w. \text{ people}}} \\ = \frac{MS_{\text{between people}} - MS_{w. \text{ people}}}{MS_{\text{between people}} + (k-1)MS_{w. \text{ people}}}.$$

In terms of the relations discussed in the last section,

$$r_1 = \frac{\overline{\text{COV}}}{\overline{\text{COV}} + (\bar{s}_X^2 - \overline{\text{COV}})} = \frac{\overline{\text{COV}}}{\bar{s}_X^2}.$$

In words, the reliability of a single measurement is estimated by the average of intercorrelations between the measurements when a pooled estimate

of the variance is used in the denominator of each of the correlations. This latter statistic will be quite close to the mean of the actual intercorrelations when the within-column variances are approximately equal.

The reliability of the average of k measurements may be expressed in terms of the reliability of a single measurement. By straightforward algebraic rearrangement of the terms in (4), it may be shown that

$$(7) \quad r_k = \frac{kr_1}{1 + (k-1)r_1}.$$

Conversely, by substitution of the expression for r_1 in (7), the latter will assume the form given in (4). Given the value of r_1 , (7) may be used to

Table 4.5-3 Numerical Example

Person	Judge 1	Judge 2	Judge 3	Judge 4	Total
(i) 1	2	4	3	3	12 = P_1
2	5	7	5	6	23 = P_2
3	1	3	1	2	7 = P_3
4	7	9	9	8	33 = P_4
5	2	4	6	1	13 = P_5
6	6	8	8	4	26 = P_6
Total	23 T_1	35 T_2	32 T_3	24 T_4	114 = G

$$(ii) (1) = \frac{G^2}{kn} = 541.50 \quad (2) = \Sigma(\Sigma X^2) = 700 \quad (3) = \frac{\Sigma T_j^2}{n} = 559.00 \quad (4) = \frac{\Sigma P_i^2}{k} = 664.00$$

	SS _{between people}	= (4) - (1)	= 122.50
	SS _{w. people}	= (2) - (4)	= 36.00
(iii)	SS _{between judges}	= (3) - (1)	= 17.50
	SS _{res}	= (2) - (3) - (4) + (1)	= 18.50
	SS _{total}	= (2) - (1)	= 158.50

estimate the reliability of the mean of any number of comparable measures. In the area of psychometrics (7) is known as the Spearman-Brown prediction formula. The assumptions underlying the validity of this formula are those underlying the analysis of variance given in Table 4.5-2. These assumptions are that the error of measurement is uncorrelated with the true score, that the sample of n people on whom the observations are made is a random sample from the population to which inferences are to be made, that the sample of k measuring instruments used is a random sample from a population of comparable instruments, and that the within-person variance may be pooled to provide an estimate of σ_{η}^2 .

The data in Table 4.5-3 will be used to illustrate the numerical application of reliability estimation. In this example, four judges were asked to rate each member of a group of six people on a specified characteristic. The

data in the columns of part i represent the ratings made by each of the judges. The data in the rows are the four ratings for a single person. For example, the ratings on person 1 are respectively 2, 4, 3, and 3. The totals of the ratings on each person are given under the column headed Total. In the numerical work, the judges have the role of the treatments.

In part ii the computational symbols used to obtain the sums of squares are shown. These symbols have the same definitions as those given in Sec. 4.2. The sums of squares are computed in part iii. Only between- and within-people sums of squares are required to estimate the reliability as defined by (4). The additional sums of squares that are computed will be discussed in connection with an alternative definition of reliability. The analysis of variance is summarized in Table 4.5-4.

Table 4.5-4 Analysis of Variance

Source of variation	SS	df	MS
Between people	122.50	5	24.50
Within people	36.00	18	2.00
Between judges	17.50	3	5.83
Residual	18.50	15	1.23
Total	158.50	23	

The estimate of reliability of the average of the four ratings made on each of the people, based upon (4), is

$$r_4 = 1 - \frac{MS_{w. \text{ people}}}{MS_{\text{between people}}} = 1 - \frac{2.00}{24.50} = .92.$$

One interpretation of this reliability is as follows: If the experiment were to be repeated with another random sample of four judges, but with the same people, the correlation between the mean ratings obtained from the two sets of data on the same people would be approximately .92. This interpretation assumes that the variance due to differences between the mean ratings made by the judges is part of the error of measurement and does not represent a systematic source of variation.

On the other hand if the mean of the ratings made by individual judges represents a systematic frame of reference for the individual judges, then the source of variation due to differences between these means should not be considered part of the error of measurement. Adjustments for differences in frame of reference may be computed as follows:

	Judge 1	Judge 2	Judge 3	Judge 4	
Total	23	35	32	24	114
Mean	3.83	5.83	5.33	4.00	4.75 = \bar{G}
Deviation from \bar{G}	-.92	1.08	.59	-.75	

The deviation of the mean rating of a judge from the mean rating of all judges defines what may be called an adjustment for the frame of reference. If, for example, the quantity 1.08 is subtracted from all the ratings made by judge 2, the mean of the resulting adjusted ratings will be equal to the grand mean. Similarly, if the quantity $-.92$ is subtracted from all ratings made by judge 1, the mean of his adjusted ratings will be equal to the grand mean. For data adjusted in this way the within-people variation is free of any source of variation which is a function of differences in frames of reference for the judges.

Table 4.5-5 Numerical Example (Adjusted Data)

Person	Judge 1	Judge 2	Judge 3	Judge 4	Total
1	2.92	2.92	2.41	3.75	12
2	5.92	5.92	4.41	6.75	23
3	1.92	1.92	.41	2.75	7
4	7.92	7.92	8.41	8.75	33
5	2.92	2.92	5.41	1.75	13
6	6.92	6.92	7.41	4.75	26
Total	28.52	28.52	28.46	28.50	114

(ii) $(1) = \frac{G_2}{kn} = 541.50$ $(2) = \sum \sum X^2 = 682.50$ $(3) = \frac{\sum T_j^2}{n} = 541.50$
 $(4) = \frac{\sum P_j^2}{k} = 664.00$

(iii) $SS_{\text{between people}} = (4) - (1) = 122.50$
 $SS_{w, \text{ people}} = (2) - (4) = 18.50$
 $SS_{\text{between judges}} = (3) - (1) = .00$
 $SS_{\text{res}} = (2) - (3) - (4) + (1) = 18.50$
 $SS_{\text{total}} = (2) - (1) = 141.00$

The data in Table 4.5-3, adjusted for differences in frames of reference, are given in Table 4.5-5. The entries under Judge 1 are obtained by subtracting the quantity $-.92$ from the corresponding entry under Judge 1 in Table 4.5-3. The entries under Judge 3 are obtained by subtracting .59 from the corresponding observed data. The entries in part i of Table 4.5-5 represent the ratings adjusted for frames of reference. The sum of the adjusted ratings for each of the people rated is the same as the corresponding sum for the observed ratings. The column totals for the adjusted ratings are, except for rounding error, all equal. Thus for the adjusted ratings there are no differences between the mean ratings given by each judge. The computational symbols associated with the sums of squares are given in part ii, and the sums of squares are given in part iii. The $SS_{\text{between people}}$ is not affected by the adjustment procedure. However, $SS_{w, \text{ people}}$ for the

adjusted data differs from the corresponding sums of squares for the unadjusted data by an amount equal to SS_{judges} . Thus the effect of the adjustment for frames of reference is to eliminate the between-judge variation from the within-person variation.

The analysis of the adjusted data is given in Table 4.5-6. Since the variation due to the judges is eliminated from the within-person variation in the course of the adjustment process, the degrees of freedom for this source of variation are also removed from the degrees of freedom for the within-person variation. Hence the degrees of freedom for the adjusted within-person variation are $18 - 3 = 15$. It should be noted that the $MS_{w. \text{ people}}$

Table 4.5-6 Analysis of Variance (Adjusted Data)

Source of variation	SS	df	MS
Between people	122.50	5	24.50
Within people	18.50	15'	1.23
Total	141.00	20'	

for the adjusted data is numerically equal to the MS_{res} for the unadjusted data. In terms of the adjusted analysis of variance, the reliability of the mean of four ratings is, by using (4),

$$r_4 = \frac{MS_{\text{between people}} - MS_{w. \text{ people (adj)}}}{MS_{\text{between people}}} \\ = \frac{24.50 - 1.23}{24.50} = .95.$$

This estimate of reliability could have been obtained from Table 4.5-4 by using MS_{res} in place of $MS_{w. \text{ people}}$ in (4). The estimate of the reliability of a single rating for the adjusted data, by using (6), is

$$r_1 = \frac{24.50 - 1.23}{24.50 + (4 - 1)(1.23)} = .83.$$

The reliability of a single rating for the adjusted data is approximately equal to the average intercorrelation between ratings given by pairs of judges. The intercorrelation between pairs of judges is not changed in the adjustment process. The actual intercorrelations are:

	Judge 1	Judge 2	Judge 3	Judge 4
Judge 1		1.00	.87	.87
Judge 2			.87	.87
Judge 3				.59

The average of these correlations is .84.

By means of (7) the reliability of the mean of four ratings may be obtained from the reliability of a single rating. Thus,

$$r_4 = \frac{4(.83)}{1 + (4 - 1)(.83)} = .95.$$

Whether the equivalent of the variation between judges is to be considered part of the error of measurement or whether this source of variation is to be considered systematic variation that is not part of the error of measurement depends upon the use that is to be made of the resulting data. If the zero points on the scales of measurement are of no importance, then differences due to the equivalent of frames of reference should not be considered part of the error of measurement. In this case the estimates of the reliability are given by

$$(4') \quad r_k = \frac{MS_{\text{between people}} - MS_{\text{res}}}{MS_{\text{between people}}},$$

$$(6') \quad r_1 = \frac{MS_{\text{between people}} - MS_{\text{res}}}{MS_{\text{between people}} + (k - 1)MS_{\text{res}}}.$$

Table 4.5-7 represents data that would be obtained from the administration of a test of k items. The totals P_1, P_2, \dots, P_n represent the test scores

Table 4.5-7 Representation of Test Scores

Person	Item 1	Item 2	...	Item k	Test score
1	X_{11}	X_{12}		X_{1k}	P_1
2	X_{21}	X_{22}		X_{2k}	P_2
.
.
.
n	X_{n1}	X_{n2}		X_{nk}	P_n
	T_1	T_2	...	T_k	G

for the people taking the tests, provided that the score on the test is obtained by summing the scores on the individual items. An estimate of the reliability of the test may be obtained from (4'). The reliability of a total is the same as that of a mean score. The computational procedures used in Sec. 4.2, with the test items having the role of the treatments, will provide the quantities required in (4'). If the test items are scored 1 for correct and 0 for incorrect, the only information required to obtain an estimate of the reliability of the test is the scores on the tests and the number of correct responses to each item. (In cases where the observed data consist of 0's and 1's, $\Sigma X = \Sigma X^2$.) The error of measurement in this case may be interpreted as a measure of the extent to which people having the same test scores do not have identical profiles of correct responses. For person 1, the profile of responses is $X_{11}, X_{12}, \dots, X_{1k}$.

One additional point about the model underlying reliability should be made explicit and emphasized. Just as in the basic model for correlated observations discussed in Sec. 4.4, one must assume that the π_i 's are constant under all measurements made. This assumption implies that the correlation between the judges (tests, items) be constant. In particular the analysis-of-variance model cannot be used to estimate reliability when the true score changes irregularly from one measurement to the next, as, for example, when practice effects are present in some nonsystematic manner. If, however, changes in the underlying true score are systematic and constant for all subjects, then adjustments for this change may be made by eliminating variation due to change from the within-subject variation.

4.6 Tests for Trend

The methods of curve fitting discussed in Sec. 3.7 may be adapted for use in connection with experiments involving repeated measures. These methods will be illustrated by means of a learning experiment.

Table 4.6-1 Analysis of Learning Data

Source of variation	SS	df	MS	F
Between subjects	90.00	9	10.00	
Within subjects	163.88	60		
Blocks	103.94	6	17.32	18.23
Residual	51.30	54	.95	

A sample of 10 subjects is used in the experiment to be described. Each subject is given 28 trials in which to learn a discrimination problem. The trials are grouped into blocks of 4 trials each. The block is considered as the observational unit. Hence the data are analyzed as if there were 7 observations on each subject, an observation being the outcome of a series of 4 trials. The degrees of freedom for the between-subject variation are 9; the degrees of freedom for the within-subject variation are 60, 6 degrees of freedom for each of the 10 subjects.

A summary of the over-all analysis appears in Table 4.6-1. For $\alpha = .01$ the critical value for a test of the hypothesis that there is no difference in the block means is $F_{.99}(1,54) = 7.14$. The data indicate that there are significant differences between the block means. A graph of the block totals is given in Fig. 4.2. Inspection of this figure indicates that a straight line would provide a good fit to the points, but there is also some evidence to indicate that an S-shaped (cubic) curve would provide a better fit. The nature of the subject matter also suggests that a cubic curve would be more appropriate for these data. Hence a point to be investigated is whether or not a cubic curve provides a better fit than a straight line (within the range of blocks included in this study).

Toward this end, the mean squares corresponding to the linear and cubic trends are computed in Table 4.6-2. The block totals for these experimental data are given near the top of Table 4.6-2. Each of these totals is the sum

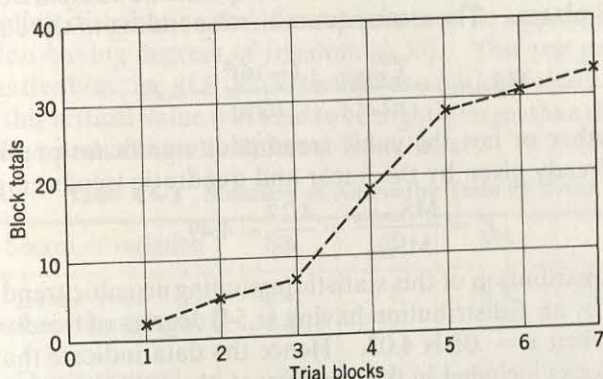


Figure 4.2

over 10 observations. Since there are seven blocks, the coefficients corresponding to the linear, quadratic, and cubic trends are obtained from the set of coefficients for which $k = 7$ in Table B.10. The entries under the column headed Σc^2 represent the sums of the squares of the coefficients in the corresponding rows. The entries under the column headed C represent the

Table 4.6-2 Tests for Trends

	Blocks:	1	2	3	4	5	6	7	Σc^2	C	MS
	Block totals:	2	5	7	18	28	31	33			
Linear		-3	-2	-1	0	1	2	3	28	166	98.41**
Quadratic		5	0	-3	-4	-3	0	5	84	-2	.00
Cubic		-1	1	1	0	-1	-1	1	6	-16	4.27*
											102.68

numerical value of the comparisons. For example, the linear comparison is $(-3)(2) + (-2)(5) + (-1)(7) + (0)(18) + (1)(28) + (2)(31) + (3)(33) = 166$.

The mean square corresponding to the linear comparison is

$$MS_{lin} = \frac{C_{lin}^2}{n\Sigma c^2} = \frac{(166)^2}{10(28)} = 98.41.$$

A test on the significance of the linear trend is given by the F ratio

$$F = \frac{MS_{lin}}{MS_{res}} = \frac{98.41}{.95} = 103.59.$$

The sampling distribution of this statistic (assuming no linear trend) may be approximated by an F distribution having degrees of freedom (1,54). The linear trend is significant beyond the .01 level.

For these data the numerical value of the quadratic comparison is zero to two decimal places. The mean square corresponding to the cubic comparison is

$$MS_{\text{cubic}} = \frac{C_{\text{cubic}}^2}{n\Sigma c^2} = \frac{(-16)^2}{10(6)} = 4.27.$$

A test on whether or not the cubic trend adds significant predictability beyond that already given by the linear and quadratic trends employs the statistic

$$F = \frac{MS_{\text{cubic}}}{MS_{\text{res}}} = \frac{4.27}{.95} = 4.49.$$

The sampling distribution of this statistic (assuming no cubic trend) may be approximated by an F distribution having (1,54) degrees of freedom. The critical value when $\alpha = .05$ is 4.03. Hence the data indicate that, within the range of blocks included in the experiment, the cubic comparison does add significant predictability to that given by the linear trend ($\alpha = .05$).

The total variation between blocks (as given in Table 4.6-1) is 103.94. Of this total, 102.68 is accounted for by the linear and cubic trends. The remaining between-block variation appears negligible relative to experimental error. The between-block variation due to higher-order trend components is

$$\begin{aligned} SS_{\text{higher order}} &= SS_{\text{blocks}} - SS_{\text{lin}} - SS_{\text{quad}} - SS_{\text{cubic}} \\ &= 1.23. \end{aligned}$$

The corresponding mean square is

$$MS_{\text{higher order}} = \frac{1.23}{3} = .41.$$

The F ratio in the test for trend components higher than the third is

$$F = \frac{MS_{\text{higher order}}}{MS_{\text{res}}} = \frac{.41}{.95} < 1.$$

Since this ratio is less than unity, the data indicate that no component higher than the third is relevant.

An alternative, more widely used testing procedure for trends is possible. In this procedure the error term in the test for linear trend is obtained from

$$\begin{aligned} SS_{\text{dev lin}} &= SS_{\text{res}} + (SS_{\text{blocks}} - SS_{\text{lin}}) \\ &= 51.30 + (103.94 - 98.41) = 56.83. \end{aligned}$$

The degrees of freedom for $SS_{\text{dev lin}}$ in this type of design are $n(k-1)-1$. For the case under consideration, the degrees of freedom are $60-1=59$. The mean square for deviations from linearity is

$$MS_{\text{dev lin}} = \frac{SS_{\text{dev lin}}}{59} = .96.$$

Under this testing procedure the statistic used in the test for linear trend is

$$F = \frac{MS_{\text{lin}}}{MS_{\text{dev lin}}} = \frac{98.41}{.96} = 102.51.$$

The sampling distribution for this statistic may be approximated by an F distribution having degrees of freedom (1,59). This test procedure has a slight negative bias; i.e., if $F_{.95}(1,59)$ is used as a critical value for a test having $\alpha = .05$, this critical value will tend to be slightly larger than the exact critical value if the trend is actually different from linear.

Table 4.6-3 Summary of Alternative Tests for Trend

Source of variation	SS	df	MS	F
Linear trend	98.41	1	98.41	102.51**
Dev from lin	56.83	59	.96	
Quadratic trend	.00	1		
Dev from quad	56.83	58		
Cubic trend	4.27	1	4.27	4.64*
Dev from cubic	52.56	57	.92	

Under the alternative test procedure, the test for cubic trend uses a denominator obtained from

$$\begin{aligned} SS_{\text{dev cubic}} &= SS_{\text{res}} + (SS_{\text{blocks}} - SS_{\text{lin}} - SS_{\text{quad}} - SS_{\text{cubic}}) \\ &= 51.30 + (103.94 - 98.41 - .00 - 4.27) = 52.56. \end{aligned}$$

The degrees of freedom for this source of variation are $n(k-1) - 3$, which in this case is $60 - 3 = 57$. The mean square corresponding to this source of variation is

$$MS_{\text{dev cubic}} = \frac{SS_{\text{dev cubic}}}{57} = \frac{52.56}{57} = .92.$$

The test for cubic trend uses the statistic

$$F = \frac{MS_{\text{cubic}}}{MS_{\text{dev cubic}}} = \frac{4.27}{.92} = 4.64.$$

The sampling distribution of this statistic may be approximated by an F distribution having (1,57) degrees of freedom. The critical value for this test ($\alpha = .05$) is 4.01. Hence the data contradict the hypothesis that the cubic trend adds no predictability to the linear and quadratic trends.

From some points of view this alternative test procedure is to be preferred to that which has a constant denominator. Under this procedure, the denominator of F ratios tends to err on the side of being too large. When the degrees of freedom for MS_{res} are large (say, over 30), the first approach presented differs only slightly from the alternative approach.

4.7 Analysis of Variance for Ranked Data

Suppose that an experimenter is interested in determining whether or not there is any difference between various methods of packaging a product. In an experiment, subjects are asked to rank the methods in order of preference. A numerical example of this kind of experiment is given in Table 4.7-1. Person 1, for example, assigned rank 1 to method 3, rank 2 to method 2, rank 3 to method 1, and rank 4 to method 4.

Table 4.7-1 Numerical Example

Person	Method 1	Method 2	Method 3	Method 4	Total
1	3	2	1	4	10
2	4	3	1	2	10
3	2	4	1	3	10
4	1	3	2	4	10
5	2	3	1	4	10
6	1	4	2	3	10
7	2	3	1	4	10
8	1	4	2	3	10
Total	16	26	11	27	80

$$(1) = \frac{G^2}{kn} = 200.00 \quad (2) = \sum \sum X^2 = 240 \quad (3) = \frac{\sum T_i^2}{n} = \frac{1782}{8} = 222.75$$

$$(4) = \frac{\sum P_i^2}{k} = \frac{800}{4} = 200.00$$

$$SS_{\text{methods}} = (3) - (1) = 22.75$$

$$SS_{\text{res}} = (2) - (3) - (4) + (1) = 17.25$$

$$SS_{\text{w. people}} = (2) - (4) = 40.00$$

The computational formulas in Sec. 4.2 may be used to obtain sums of squares required for the analysis of variance. $SS_{\text{between people}}$ will always be zero. Upon replicating the experiment with random samples of subjects, SS_{total} will remain constant provided that no tied ranks are permitted. Rather than an F statistic, the chi-square statistic

$$(1) \quad \chi_{\text{ranks}}^2 = \frac{SS_{\text{methods}}}{(SS_{\text{methods}} + SS_{\text{res}})/n(k-1)}$$

$$= \frac{n(k-1)SS_{\text{methods}}}{SS_{\text{w. people}}}$$

is used to test the hypothesis of no difference in mean rank for the methods. The higher the agreement between people in ranking the methods, the larger SS_{methods} will be. For the data in Table 4.7-1,

$$\chi_{\text{ranks}}^2 = \frac{8(3)(22.75)}{40.00} = 13.65.$$

The critical value of this statistic for a .01-level test is

$$\chi_{.99}^2(k-1) = \chi_{.99}^2(3) = 11.3.$$

Since the observed χ^2 exceeds the critical value, the data contradict the hypothesis of no difference between the mean ranks for the different methods of packaging. Inspection of Table 4.7-1 indicates that methods 1 and 3 have the smaller means and methods 2 and 4 the larger means.

When no tied ranks are permitted,

$$SS_{w. \text{ people}} = \frac{nk(k^2 - 1)}{12},$$

and the χ^2 statistic becomes

$$(2) \quad \chi_{\text{ranks}}^2 = \frac{12}{nk(k+1)} (\sum T_j^2) - 3n(k+1).$$

The expression for χ^2 in (2) is algebraically equivalent to (1) when no ties are permitted; (1) may be used whether or not ties are present. For the data in Table 4.7-1, expression (2) gives

$$\begin{aligned} \chi_{\text{ranks}}^2 &= \frac{12}{8(4)(5)} (1782) - 3(8)(5) \\ &= 13.65. \end{aligned}$$

Use of this statistic in testing the hypothesis on the mean ranks defines what is called the Friedman test.

An index of the extent to which people agree in their preferences is given by the coefficient of concordance, which is defined as

$$(3) \quad W = \frac{SS_{\text{methods}}}{SS_{\text{total}}}.$$

This coefficient is related to the average intercorrelation between the rankings assigned by the people; this relationship is

$$(4) \quad \bar{r} = \frac{nW - 1}{n - 1}.$$

The test in (1) may be regarded as a test of the hypothesis that the coefficient of concordance in the population of people is zero. For the data in Table 4.7-1,

$$W = \frac{22.75}{40.00} = .569.$$

The average intercorrelation between the people is

$$\bar{r} = \frac{8(.569) - 1}{8 - 1} = .507.$$

The index W corresponds to a correlation ratio, whereas \bar{r} corresponds to a product-moment correlation (or equivalently a rank-difference correlation).

The average intercorrelation between the people may also be computed by means of relationship (6') in Sec. 4.5, provided that n is used in place of k . In terms of this relationship,

$$\begin{aligned}\bar{r} &= \frac{MS_{\text{between methods}} - MS_{\text{res}}}{MS_{\text{between methods}} + (n - 1)MS_{\text{res}}} \\ &= \frac{7.583 - .821}{7.583 + 7(.821)} = .507.\end{aligned}$$

4.8 Dichotomous Data

Observed data may in some cases be classified into one of two classes; for example, a characteristic may be present or absent, a response is either yes or no, a drug is either fatal or it is not. Such data are said to be dichotomous. One of the two dichotomies may conveniently be designated by a 0, the other by a 1. Which category is assigned the zero is arbitrary.

Table 4.8-1 Numerical Example

Subject	Time 1	Time 2	Time 3	Time 4	Time 5	Total
1	0	0	0	0	0	0 = P_1
2	0	0	1	1	0	2 = P_2
3	0	0	1	1	1	3
4	0	1	1	1	1	4
5	0	0	0	0	1	1
6	0	1	0	1	1	3
7	0	0	1	1	1	3
8	1	0	0	1	1	3
9	1	1	1	1	1	5
10	1	1	1	1	1	5
Total	3	4	6	8	8	29

$$(1) = \frac{G^2}{kn} = \frac{29^2}{50} = 16.82 \quad (2) = \sum \sum X^2 = 29 \quad (3) = \frac{\sum T_j^2}{n} = \frac{189}{10} = 18.90$$

$$(4) = \frac{\sum P_i^2}{k} = \frac{107}{5} = 21.40$$

$SS_{\text{between people}}$	$= (4) - (1)$	$= 4.58$	df	
$SS_{\text{w. people}}$	$= (2) - (4)$	$= 7.60$		
SS_{time}	$= (3) - (1)$	$= 2.08$	4	$MS_{\text{time}} = .520$
SS_{res}	$= (2) - (3) - (4) + (1)$	$= 5.52$	36	$MS_{\text{res}} = .153$
SS_{total}	$= (2) - (1)$	$= 12.18$		

Consider an experiment designed to study the effects of an advertising campaign upon the attitude of a potential population of buyers toward the product. The data may take the form given in Table 4.8-1. In this table a 0 is used to indicate an unfavorable attitude, and a 1 is used to indicate a favorable attitude. Suppose that a random sample of 10 subjects is selected

for the study. Each subject is interviewed at the end of five different time periods. For example, subject 1 was not favorable at any time; subject 5 was not favorable on the first four interviews but was favorable on the fifth interview.

To test the hypothesis of no change in the percentage of favorable replies over the time periods, the statistic

$$(1) \quad Q = \frac{n(k-1)SS_{\text{time}}}{SS_{\text{w, people}}}$$

may be used. The Q statistic has the same form as χ^2_{ranks} , which was used in Sec. 4.7. Under the hypothesis of no change in the percentage of favorable replies, Cochran (1950) has shown that the sampling distribution of the Q statistic is approximated by a chi-square distribution with $k-1$ degrees of freedom, when n is reasonably large. For an α -level test the critical value for the Q statistic is $\chi^2_{1-\alpha}(k-1)$.

For the data in Table 4.8-1,

$$Q = \frac{10(5-1)(2.08)}{7.60} = 10.95.$$

The critical value for a .05-level test is $\chi^2_{.95}(5-1) = 9.5$. Hence the experimental data contradict the hypothesis of no change in the percentage of favorable replies. Examination of the data indicates a systematic increase in the percentage of favorable replies. Cochran (1950) has indicated that the F statistic computed by treating the data as if the measurements were normally distributed variables will yield probability statements which are relatively close to those obtained by use of the Q statistic. For the data in Table 4.8-1,

$$F = \frac{MS_{\text{time}}}{MS_{\text{res}}} = \frac{.520}{.153} = 3.40.$$

For a .05-level test the critical value for the F statistic is $F_{.95}(4,36) = 2.63$. Thus use of the F statistic also leads to the rejection of the hypothesis of no change in the percentage of favorable replies.

CHAPTER 5

Design and Analysis of Factorial Experiments

5.1 General Purpose

Factorial experiments permit the experimenter to evaluate the combined effect of two or more experimental variables when used simultaneously. Information obtained from factorial experiments is more complete than that obtained from a series of single-factor experiments, in the sense that factorial experiments permit the evaluation of *interaction* effects. An interaction effect is an effect attributable to the combination of variables above and beyond that which can be predicted from the variables considered singly.

For example, many of the properties of the chemical substance H_2O (water) cannot be predicted from the properties of oxygen and the properties of hydrogen studied in isolation. Most of the properties of water are attributable to the effect of the interaction between oxygen and hydrogen. The compound formed by this interaction has properties which are not given by simply adding the properties of oxygen to the properties of hydrogen.

At the end of a factorial experiment, the experimenter has information which permits him to make decisions which have a broad range of applicability. In addition to information about how the experimental variables operate in relative isolation, the experimenter can predict what will happen when two or more variables are used in combination. Apart from the information about interactions, the estimates of the effects of the individual variables is, in a sense, a more practically useful one; these estimates are obtained by averaging over a relatively broad range of other relevant experimental variables. By contrast, in a single-factor experiment some relevant experimental variables may be held constant, while others may be randomized. In the case of a factorial experiment, the population to which inferences can be made is more inclusive than the corresponding population for a single-factor experiment.

In working with factorial experiments in the behavioral science area,

a sharp distinction must be drawn between experiments involving repeated measures on the same elements and those which do not involve repeated measures. The material in this chapter will be concerned primarily with experiments which do not involve repeated measures. However, many of the basic principles to be developed in this chapter will be applicable, with only slight modification, to the case in which there are repeated measures.

The term *factor* will be used in a broad sense. For some purposes a distinction will be made between treatment and classification factors. The latter group the experimental units into classes which are homogeneous with respect to what is being classified. In contrast, treatment factors define the experimental conditions applied to an experimental unit. The administration of the treatment factors is under the direct control of the experimenter, whereas classification factors are not, in a sense. The effects of the treatment factors are of primary interest to the experimenter, whereas classification factors are included in an experiment to reduce experimental error and clarify interpretation of the effects of the treatment factors.

The design of factorial experiments is concerned with answering the following questions:

1. What factors should be included?
2. How many levels of each factor should be included?
3. How should the levels of the factors be spaced?
4. How should the experimental units be selected?
5. How many experimental units should be selected for each treatment combination?
6. What steps should be taken to control experimental error?
7. What criterion measures should be used to evaluate the effects of the treatment factors?
8. Can the effects of primary interest be estimated adequately from the experimental data that will be obtained?

The answers to these questions will be considered in some detail in this chapter and the chapters that follow. An excellent and readable overview of the planning of factorial experiments will be found in Cox (1958, chap. 7).

5.2 Terminology and Notation

The term *factor* will be used interchangeably with the terms *treatment* and *experimental variable*. More specifically, a factor is a series of related treatments or related classifications. The related treatments making up a factor constitute the *levels* of that factor. For example, a factor *color* may consist of three levels: red, green, and yellow. A factor *size* may consist of two levels: small and large. A factor *dosage* may consist of four levels: 1 cc, 3 cc, 5 cc, and 7 cc. The number of levels within a factor is determined largely by the thoroughness with which an experimenter desires to investigate the factor. Alternatively, the levels of a factor are determined by the kind of inferences the experimenter desires to make upon conclusion of the

experiment. The levels of a factor may be quantitative variations in an essentially quantitative variable, or they may be qualitatively different categories within an essentially qualitative variable. Basically a factor is a qualitative variable; in special cases it becomes a quantitative variable.

The *dimensions* of a factorial experiment are indicated by the number of factors and the number of levels of each factor. For example, a factorial experiment in which there are two factors, one having three levels and the other having four levels, is called a 3×4 (read "three by four") factorial experiment. In a $2 \times 3 \times 5$ factorial experiment, there are three factors, having respective levels of two, three, and five. (Many different designs may be constructed for a given factorial experiment.) The treatment combinations in a 2×3 factorial experiment may be represented schematically as follows:

		Levels of factor <i>B</i>		
		b_1	b_2	b_3
Levels of factor <i>A</i>	a_1	ab_{11}	ab_{12}	ab_{13}
	a_2	ab_{21}	ab_{22}	ab_{23}

In this schema, a_1 and a_2 designate the levels of factor *A*; b_1 , b_2 , and b_3 designate the levels of factor *B*. In a 2×3 factorial experiment six possible combinations of treatments may be formed. Level a_1 may be used in combination with each of the three levels of factor *B*; level a_2 may also be used in combination with each of the three levels of factor *B*. The resulting treatment combinations are labeled in the cells of the schema. For example, the symbol ab_{12} represents the experimental condition resulting when factor *A* is at level a_1 and factor *B* is at level b_2 .

For the case of a $p \times q$ factorial experiment, pq different treatment combinations are possible. In a $p \times q \times r$ factorial experiment there are pqr treatment combinations. As the number of factors increases, or as the number of levels within a factor increases, the number of treatment combinations in a factorial experiment increases quite rapidly. For example, a $5 \times 5 \times 5$ factorial experiment has 125 treatment combinations.

In an experiment, the elements observed under each of the treatment combinations will generally be a random sample from some specified population. This specified population will, in most cases of interest, contain a potentially infinite number of elements. If n elements are to be observed under each treatment combination in a $p \times q$ factorial experiment, a random sample of npq elements from the population is required (assuming no repeated measurements on the same elements). The npq elements are then subdivided at random into pq subsamples of size n each. These subsamples are then assigned at random to the treatment combinations.

The potential (or population) levels of factor *A* will be designated by the symbols $a_1, a_2, \dots, a_I, \dots, a_p$. The number of such potential levels, P ,

may be quite large. The experimenter may group the P potential levels into p levels ($p < P$) by either combining adjoining levels or deliberately selecting what are considered to be representative levels. For example, if factor A represents the dimension of age, the experimenter may choose to group the levels into 1-year intervals; alternatively, the experimenter may choose to group the levels into 2-year intervals. On the other hand, if factor A represents a dosage dimension, the experimenter may deliberately select a series of representative dosages; i.e., in terms of previous research the levels selected may be representative of low, middle, and high dosages.

When p , the number of levels of factor A included in the experiment, is equal to P , then factor A is called a *fixed* factor. When the selection of the p levels from the potential P levels is determined by some systematic, non-random procedure, then factor A is also considered a fixed factor. In this latter case, the selection procedure reduces the potential P levels to p *effective* levels. Under this type of selection procedure, the effective, potential number of levels of factor A in the population may be designated $P_{\text{effective}}$, and $P_{\text{effective}} = p$.

In contrast to this systematic selection procedure, if the p levels of factor A included in the experiment represent a random sample from the potential P levels, then factor A is considered to be a *random* factor. For example, in an experiment designed to test the effectiveness of various drugs upon categories of patients, the factor A may be the hospitals in which the patients are located. Potentially the number of different hospitals may be quite large. If a random sample of p of the P potential hospitals are included in the experiment, then factor A is a random factor. (If, further, the sampling within each of the hospitals selected is proportional to the number of patients within the hospital, then conclusions drawn from this kind of experiment will be relevant to the domain of all patients and not just to the domain of all hospitals.) In most practical situations in which random factors are encountered, p is quite small relative to P , and the ratio p/P is quite close to zero.

Similar definitions apply to factor B . Let the potential levels of factor B be designated $b_1, b_2, \dots, b_j, \dots, b_Q$. Of these Q potential levels, let the number of levels actually in an experiment be q . If $q = Q$, or if the effective number of levels of factor B is reduced from Q to q by some systematic, nonrandom procedure, then factor B is considered fixed. (The reduction from Q potential levels to q effective levels is a function of the experimental design.) The actual levels of factor B included in an experiment will be designated by the notation $b_1, b_2, \dots, b_j, \dots, b_q$. If the q levels in an experiment are a random sample from the Q potential levels, then factor B is considered a random factor. In most practical cases in which factor B is a random factor, Q will be quite large relative to q , and the ratio q/Q will be close to zero.

The ratio of the number of levels of a factor in an experiment to the

potential number of levels in the population is called the *sampling fraction* for a factor. In terms of this sampling fraction, the definitions of fixed and random factors may be summarized as follows:

Sampling fraction		Factor
$p/P = 1$	or $p/P_{\text{effective}} = 1$	A is a fixed factor
	$p/P \doteq 0$	A is a random factor
$q/Q = 1$	or $q/Q_{\text{effective}} = 1$	B is a fixed factor
	$q/Q \doteq 0$	B is a random factor

Cases in which the sampling fraction assumes a value between 0 and 1 do occur in practice. However, cases in which the sampling fraction is either 1 or very close to 0 are encountered more frequently.

The symbol a_i will be used to refer to an arbitrary level of factor A when the frame of reference is the potential P levels of factor A . The symbol a_i will be used to refer to an arbitrary level of factor A when the frame of reference is the p levels actually included in the experiment. A similar distinction will be made between the symbols b_j and b_j . The first symbol has as its frame of reference the potential Q levels of factor B , whereas the second symbol has as its frame of reference the q levels of factor B in the actual experiment. In those cases in which factor A is a random factor, the symbol a_1 , when the frame of reference is the potential population levels, refers to a specified population level. However, when the frame of reference is the experiment, the symbol a_1 refers to a specified level in the experiment. Thus the symbol a_1 may refer to different levels of the same factor, depending upon the frame of reference. The context in which the notation is used will clarify the ambiguity of the symbol considered in isolation.

To illustrate additional notation, assume that all the potential elements are included in all the potential cells of a factorial experiment. (Generally only a random sample of the potential elements are included in a cell of a factorial experiment.) Assume further that, after the treatment combinations have been administered, each of the elements is measured (observed) on the characteristic being studied (the *criterion* or dependent variable). The mean of the observations made under each treatment combination will be denoted by the following notation:

	b_1	b_2	\cdots	b_J	\cdots	b_Q
a_1	μ_{11}	μ_{12}	\cdots	μ_{1J}	\cdots	μ_{1Q}
a_2	μ_{21}	μ_{22}	\cdots	μ_{2J}	\cdots	μ_{2Q}
\vdots						
a_I	μ_{I1}	μ_{I2}	\cdots	μ_{IJ}	\cdots	μ_{IQ}
\vdots						
a_P	μ_{P1}	μ_{P2}	\cdots	μ_{PJ}	\cdots	μ_{PQ}

In this notation, μ_{IJ} denotes the mean on the criterion for the potential population of elements under treatment combination ab_{IJ} . Equivalently, μ_{IJ} represents the population mean for the dependent variable in cell ab_{IJ} . It will be assumed that the potential number of elements in each of the cells is the same for all cells.

The average of the cell means appearing in row I is

$$\mu_{I.} = \frac{\sum_J \mu_{IJ}}{Q}.$$

In words, $\mu_{I.}$ is the mean of the dependent variable averaged over all potential treatment combinations in which factor A is at level a_I . Similarly, the mean of the dependent variable averaged over all potential treatment combinations in which factor B is at level b_J is

$$\mu_{.J} = \frac{\sum_I \mu_{IJ}}{P}.$$

The grand mean on the dependent variable is

$$\mu_{..} = \frac{\sum_I \sum_J \mu_{IJ}}{PQ} = \frac{\sum_I \mu_{I.}}{P} = \frac{\sum_J \mu_{.J}}{Q}.$$

The grand mean may also be defined as the mean of the dependent variable for all potential observations under all potential treatment combinations.

Notation for the statistics obtained from actual experimental data is as follows. (It is assumed that each cell mean is based upon an independent random sample of size n .)

	b_1	b_2	\cdots	b_j	\cdots	b_q	
a_1	\overline{AB}_{11}	\overline{AB}_{12}		\overline{AB}_{1j}		\overline{AB}_{1q}	\bar{A}_1
a_2	\overline{AB}_{21}	\overline{AB}_{22}		\overline{AB}_{2j}		\overline{AB}_{2q}	\bar{A}_2
\vdots							\vdots
\vdots							\vdots
\vdots							\vdots
a_i	\overline{AB}_{i1}	\overline{AB}_{i2}	\cdots	\overline{AB}_{ij}	\cdots	\overline{AB}_{iq}	\bar{A}_i
\vdots							\vdots
\vdots							\vdots
a_p	\overline{AB}_{p1}	\overline{AB}_{p2}		\overline{AB}_{pj}		\overline{AB}_{pq}	\bar{A}_p
	\bar{B}_1	\bar{B}_2	\cdots	\bar{B}_j	\cdots	\bar{B}_q	\bar{G}

The symbol \overline{AB}_{ij} represents the mean of the measurements on the dependent variable for the n elements under treatment combination ab_{ij} . The average of all observations at level a_i is

$$\bar{A}_i = \frac{\sum_j \overline{AB}_{ij}}{q}.$$

Similarly, the average of all observations at level b_j is

$$\bar{B}_j = \frac{\sum_i \bar{AB}_{ij}}{p}.$$

The grand mean \bar{G} is the mean of all means. Thus

$$\bar{G} = \frac{\sum_i \sum_j \bar{AB}_{ij}}{pq} = \frac{\sum_i \bar{A}_i}{p} = \frac{\sum_j \bar{B}_j}{q}.$$

In most designs for factorial experiments, these statistics are unbiased estimates of corresponding parameters. That is,

$$E(\bar{AB}_{ij}) = \mu_{ij}, \quad E(\bar{A}_i) = \mu_{i.}, \quad \text{and} \quad E(\bar{B}_j) = \mu_{.j}.$$

5.3 Main Effects

Main effects are defined in terms of parameters. Direct estimates of these parameters will, in most cases, be obtainable for corresponding statistics. The main effect of level a_I of factor A is by definition

$$\alpha_I = \mu_{I.} - \mu_{..}$$

In words, the main effect for level a_I is the difference between the mean of all potential observations on the dependent variable at level a_I and the grand mean of all potential observations. The main effect for level a_I may be either positive or negative. The main effect for level $a_{I'}$, where I' designates a level of factor A different from I , is

$$\alpha_{I'} = \mu_{I'.} - \mu_{..}$$

For most practical purposes, only the difference between two main effects will be needed. The *differential* main effect is defined to be

$$\alpha_I - \alpha_{I'} = \mu_{I.} - \mu_{I'..}$$

(In some experiments, only the differential main effects can be estimated from the observed data.)

Analogous definitions hold for the main effects of the levels of factor B . Thus, the main effect for level b_J is

$$\beta_J = \mu_{.J} - \mu_{..}$$

The differential main effect for levels b_J and $b_{J'}$ is

$$\beta_J - \beta_{J'} = \mu_{.J} - \mu_{.J'..}$$

A differential main effect measures the extent to which criterion means for two levels within the same factor differ. Thus, if all the means for the various levels of a factor are equal, all the differential main effects for the levels of that factor will be zero. It should be pointed out that equality of the population means does not imply equality of the sample main effects.

Hence estimates of differential main effects may not be zero even though the population values are zero. However, when the population differential main effects are zero, their sample estimates will differ from zero by amounts which are functions of experimental error.

The variance of the main effects due to factor A is, by definition,

$$\sigma_{\alpha}^2 = \frac{\sum_I (\mu_{I.} - \mu_{..})^2}{P - 1} = \frac{\sum_I \alpha_I^2}{P - 1}.$$

An equivalent definition in terms of the differential main effects is

$$\sigma_{\alpha}^2 = \frac{\sum (\alpha_I - \alpha_{I'})^2}{P(P - 1)}, \quad (I < I').$$

The symbol $I < I'$ indicates that the summation is over different pairs of α 's; that is, $\alpha_1 - \alpha_2$ is not considered to be different from $\alpha_2 - \alpha_1$. The variance of the main effects of factor A measures the extent to which the criterion means for the various levels of factor A differ. Equivalently, the variance of the main effects may be regarded as an over-all measure of the differential main effects for that factor. Thus, σ_{α}^2 will be small when all the differential main effects are small; σ_{α}^2 will be large when one or more of the differential main effects are large. When all the $\mu_{I.}$ are equal, σ_{α}^2 will be zero.

Analogous definitions hold for the variance of the main effects of factor B .

$$\sigma_{\beta}^2 = \frac{\sum_J (\mu_{.J} - \mu_{..})^2}{Q - 1} = \frac{\sum_J \beta_J^2}{Q - 1}.$$

In terms of the differential main effects,

$$\sigma_{\beta}^2 = \frac{\sum_J (\beta_J - \beta_{J'})^2}{Q(Q - 1)}, \quad (J < J').$$

To illustrate the definitions of main effects and their variance, suppose the population means on the dependent (criterion) variable are those given in the following table. Here $P = 2$ and $Q = 3$.

	b_1	b_2	b_3	Mean
a_1	10	5	15	10
a_2	20	5	5	10
Mean	15	5	10	10

For the data in this table, $\mu_{1.} = 10$, $\mu_{2.} = 10$, and $\mu_{..} = 10$. Hence $\alpha_1 = 0$, and $\alpha_2 = 0$. It is also noted that $\mu_{.1} = 15$, $\mu_{.2} = 5$, $\mu_{.3} = 10$. Hence

$$\beta_1 = 15 - 10 = 5,$$

$$\beta_2 = 5 - 10 = -5,$$

$$\beta_3 = 10 - 10 = 0.$$

The differential main effects for factor B are

$$\beta_1 - \beta_2 = 10, \quad \beta_1 - \beta_3 = 5, \quad \beta_2 - \beta_3 = -5.$$

[In the general case there will be $Q(Q - 1)/2$ distinct differential main effects for factor B .]

The variance due to the main effects of factor B is

$$\sigma_B^2 = \frac{\sum \beta_J^2}{Q - 1} = \frac{5^2 + (-5)^2 + 0^2}{3 - 1} = 25.$$

In terms of the differential main effect of factor B , the variance is

$$\begin{aligned} \sigma_B^2 &= \frac{\Sigma(\beta_J - \beta_{J'})^2}{Q(Q - 1)} \\ &= \frac{10^2 + 5^2 + (-5)^2}{3(3 - 1)} = 25. \end{aligned}$$

In summary, main effects as well as differential main effects are defined in terms of a specified population and not in terms of individual treatments within the population. Thus a main effect is *not* a parameter associated with a specified level of a specified factor; rather, a main effect depends upon all the other factors that may be present, as well as the number of levels assumed for the specified factor. It should also be noted that main effects are computed from means which are obtained by averaging over the totality of all other factors present in the population to which inferences are to be made.

5.4 Interaction Effects

The interaction between level a_I and level b_J , designated by the symbol $\alpha\beta_{IJ}$, is a measure of the extent to which the criterion mean for treatment combination ab_{IJ} cannot be predicted from the sum of the corresponding main effects. From many points of view, interaction is a measure of the nonadditivity of the main effects. To some extent the existence or non-existence of interaction depends upon the scale of measurement. For example, in terms of a logarithmic scale of measurement, interaction may not be present, whereas in terms of some other scale of measurement an interaction effect may be present. The choice of a scale of measurement for the dependent variable is generally at the discretion of the experimenter. If alternative choices are available, then that scale which leads to the simplest additive model will generally provide the most complete and adequate summary of the experimental data.

In terms of the population means and the population main effects,

$$\alpha\beta_{IJ} = \mu_{IJ} - (\alpha_I + \beta_J + \mu_{..}).$$

From the definition of the main effects, the above expression may be shown to be algebraically equivalent to

$$\alpha\beta_{IJ} = \mu_{IJ} - \mu_{I.} - \mu_{.J} + \mu_{..}$$

From the way in which $\mu_{I.}$ and $\mu_{.J}$ were defined, it follows that

$$\begin{aligned}\sum_I \alpha\beta_{IJ} &= \sum_I \mu_{IJ} - \sum_I \mu_{I.} - \sum_I \mu_{.J} + \sum_I \mu_{..} \\ &= P\mu_{.J} - P\mu_{..} - P\mu_{.J} + P\mu_{..} \\ &= 0.\end{aligned}$$

It also follows that

$$\sum_J \alpha\beta_{IJ} = 0.$$

In words, the sum of the interaction effects within any row or any column of the population cells is zero.

Differential interaction effects are defined by the following examples:

- (i) $\alpha\beta_{12} - \alpha\beta_{34} = \mu_{12} - \mu_{34} - (\alpha_1 - \alpha_3) - (\beta_2 - \beta_4),$
- (ii) $\alpha\beta_{12} - \alpha\beta_{13} = \mu_{12} - \mu_{13} - (\beta_2 - \beta_3),$
- (iii) $\alpha\beta_{12} - \alpha\beta_{32} = \mu_{12} - \mu_{32} - (\alpha_1 - \alpha_3).$

In (i) the differential interaction effect depends upon two differential main effects as well as the difference between cell means. In (ii) and (iii) only a single differential main effect is involved; (ii) depends upon the differential main effect associated with two levels of factor B , whereas (iii) depends upon a differential main effect associated with two levels of factor A . Relative to (i), the differential interaction effects represented by (ii) and (iii) are classified as *simple* interaction effects.

The variance due to interaction effects in the population is, by definition,

$$\sigma_{\alpha\beta}^2 = \frac{\sum \sum (\alpha\beta)_{IJ}^2}{(P-1)(Q-1)}.$$

An equivalent definition can also be given in terms of the differential interaction effects. Under the hypothesis that the individual cell means may be predicted from a knowledge of corresponding main effects, $\sigma_{\alpha\beta}^2 = 0$; that is, under the hypothesis of additivity of main effects, the variance due to the interaction effect is equal to zero.

For the numerical data given in the last section,

$$\begin{aligned}\alpha\beta_{11} &= 10 - 10 - 15 + 10 = -5, \\ \alpha\beta_{12} &= 5 - 10 - 5 + 10 = 0, \\ \alpha\beta_{13} &= 15 - 10 - 10 + 10 = 5.\end{aligned}$$

It is also noted that $\sum_J \alpha\beta_{1J} = -5 + 0 + 5 = 0$.

5.5 Experimental Error and Its Estimation

All uncontrolled sources of variance influencing an observation under a specified treatment combination contribute to what is known as the variance due to experimental error. The greater the number of relevant sources of variance which are controlled and measured, the smaller the variance due to experimental error. In a population in which there are N potential observations (i.e., measurements on a criterion) under a specified treatment combination, the variance of these observations defines the variance due to experimental error. Since all the observations within this cell have been made under the same treatment combination, effects associated with the treatment per se have been held constant. Hence the treatment effects per se are not contributing to the within-cell variance. Differences between units of the experimental material existing prior to the experimental treatment, variance introduced by inaccuracies or uncontrolled changes in experimental techniques, possible unique interactions between the material and the treatments—all these sources contribute to the within-cell variance.

If an observation on element K under treatment combination ab_{IJ} is designated by the symbol X_{IJK} , then the variance of the N potential elements under this treatment combination is given by

$$\sigma_{IJ}^2 = \frac{\sum_K (X_{IJK} - \mu_{IJ})^2}{N - 1}.$$

[In the usual definition of a population variance, N rather than $N - 1$ appears as the divisor. However, using $N - 1$ as the divisor simplifies the notation that follows. This definition of σ_{IJ}^2 is the one adopted by Cornfield and Tukey (1956, p. 916). As N approaches infinity, $N - 1$ approaches N .] Thus, σ_{IJ}^2 , the within-cell variance for the population cell ab_{IJ} , is the variance due to experimental error in this cell. For purposes of making tests of significance and obtaining confidence bounds on parameters, it must be assumed that the variance due to experimental error is constant for all cells in the experiment. This is the assumption of homogeneity of error variance; this assumption may be represented symbolically by

$$\sigma_{IJ}^2 = \sigma_e^2 \quad \text{for all } IJ\text{'s,}$$

where σ_e^2 is the variance due to experimental error within any cell in the population. For the same purposes, it will generally be assumed that X_{IJK} is normally distributed within each of the cells.

In an experiment, assume that there are n independent observations within each of the treatment combinations included in the experiment. The variance of the observations in the cell ab_{ij} is given by

$$s_{ij}^2 = \frac{\sum_k (X_{ijk} - \bar{AB}_{ij})^2}{n - 1}.$$

Assuming that the experimental error in the population is σ_e^2 for all cells, s_{ij}^2 provides an estimate of σ_e^2 . A better estimate of σ_e^2 is obtained by averaging the within-cell variances for each of the pq cells in the experiment. This average within-cell variance, denoted by the symbol s_{pooled}^2 , may be represented as

$$s_{\text{pooled}}^2 = \frac{\sum \sum s_{ij}^2}{pq}.$$

Under the assumptions made, s_{pooled}^2 will be an unbiased estimate of σ_e^2 . Further, the sampling distribution of $pq(n-1)s_{\text{pooled}}^2/\sigma_e^2$ may be approximated by a chi-square distribution having $pq(n-1)$ degrees of freedom. [Since the variance within each of the pq cells in the experiment is based upon n independent observations, each of the within-cell variances has $n-1$ degrees of freedom. The pooled variance, s_{pooled}^2 , has degrees of freedom equal to the combined degrees of freedom for each of the within-cell variance, that is, $pq(n-1)$.] Because the observations within each of the cells are independent, the degrees of freedom for each of the variances that are averaged are additive.

The statistic s_{pooled}^2 is called the within-cell mean square, abbreviated $\text{MS}_{\text{w. cell}}$. Alternatively, this source of variance is designated as MS_{error} in designs which do not have repeated measures.

5.6 Estimation of Mean Squares Due to Main Effects and Interaction Effects

Some of the assumptions that will be made for purposes of estimation and analysis in a two-factor experiment are summarized by the following structural model:

$$(1) \quad X_{ijk} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ijk}.$$

In this model, X_{ijk} is an observation made in the experiment on element k under treatment combination ab_{ij} . On the right-hand side of (1) are the factorial effects and the experimental error. This model assumes that the factorial effects as well as the experimental error are additive, i.e., that an observation is a linear function of the factorial effects and the experimental error. Expression (1) is a special case of the general linear hypothesis. Each of the terms on the right-hand side is assumed to be independent of the others.

The left-hand side of (1) represents an observation on the dependent variable made in the experiment. The terms on the right-hand side of (1) are parameters of the independent variables which account for the differences between observations. The terms on the right-hand side cannot be observed directly; however, data from the experiment will generally permit the experimenter to estimate certain linear functions of these effects and test hypotheses about such functions.

For all observations made under treatment combination ab_{ij} , the terms α_i , β_j , and $\alpha\beta_{ij}$ are constant. ($\mu_{..}$ is a constant for all observations.) Hence the only source of variation for observations made under treatment combination ab_{ij} is that due to experimental error. If s_{ij}^2 is the variance of the observations within cell ab_{ij} , then s_{ij}^2 is an estimate of σ_e^2 .

In terms of (1), the mean of the observations within cell ab_{ij} may be expressed as

$$(2) \quad \overline{AB}_{ij} = \frac{\sum X_{ijk}}{n} = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + \bar{\epsilon}_{ij}.$$

The notation $\bar{\epsilon}_{ij}$ denotes the mean experimental error for the n observations in cell ab_{ij} . If the observations in cell ab_{ij} were to be replicated with an independent sample of size n , the term $\bar{\epsilon}_{ij}$ would not remain constant.

Table 5.6-1 Structural Parameters Estimated by Various Statistics

Statistic	Structural parameters estimated
$\bar{A}_i - \bar{A}_{i'}$	$(\alpha_i - \alpha_{i'}) + (\bar{\alpha}\beta_i - \bar{\alpha}\beta_{i'}) + (\bar{\epsilon}_i - \bar{\epsilon}_{i'})$
$\bar{B}_j - \bar{B}_{j'}$	$(\beta_j - \beta_{j'}) + (\bar{\alpha}\beta_j - \bar{\alpha}\beta_{j'}) + (\bar{\epsilon}_j - \bar{\epsilon}_{j'})$
$\overline{AB}_{ij} - \overline{AB}_{i'j'}$	$(\alpha_i - \alpha_{i'}) + (\beta_j - \beta_{j'}) + (\alpha\beta_{ij} - \alpha\beta_{i'j'}) + (\bar{\epsilon}_{ij} - \bar{\epsilon}_{i'j'})$
$(\overline{AB}_{ij} - \bar{A}_i - \bar{B}_j) - (\overline{AB}_{i'j'} - \bar{A}_{i'} - \bar{B}_{j'})$	$(\alpha\beta_{ij} - \alpha\beta_{i'j'}) + (\bar{\epsilon}_{ij} - \bar{\epsilon}_{i'j'})$

However, all the other terms on the right-hand side of (2) would remain constant for this kind of replication. Thus, for a large number of replications of the observations within cell ab_{ij} , each replication with an independent sample of size n ,

$$(3) \quad E(\overline{AB}_{ij}) = E(\mu_{..}) + E(\alpha_i) + E(\beta_j) + E(\alpha\beta_{ij}) + E(\bar{\epsilon}_{ij}) \\ = \mu_{..} + \alpha_i + \beta_j + \alpha\beta_{ij} + 0.$$

In this context, the expected value of terms on the right-hand side is the average value of a large number of replications for cell ab_{ij} with independent samples. Although the mean experimental error $\bar{\epsilon}_{ij}$ has expected value equal to zero, it varies from one replication to the next. Hence the variance of $\bar{\epsilon}_{ij}$ is not zero.

From (1) and the definitions of the various parameters, the statistics summarized in Table 5.6-1 have the expressions indicated. The right-hand side of this table gives the structural variables estimated by the statistics on the left-hand side.

The notation $\bar{\alpha}\beta_i$ denotes the average for level a_i of the $\alpha\beta_{ij}$ effects over the levels of factor B included in the experiment. If $q = Q$ (that is, if factor B is fixed), then $\bar{\alpha}\beta_i$ is zero. If $q \neq Q$, $\bar{\alpha}\beta_i$ need not be zero for any single

experiment. To illustrate this point, suppose that there are six levels of factor B in the population of such levels, and suppose that the interaction effects associated with level a_i are those given below:

	b_1	b_2	b_3	b_4	b_5	b_6
a_i	$\alpha\beta_{i1} = 3$	$\alpha\beta_{i2} = 2$	$\alpha\beta_{i3} = -3$	$\alpha\beta_{i4} = -2$	$\alpha\beta_{i5} = 5$	$\alpha\beta_{i6} = -5$

In this case $Q = 6$. Note that $\sum_j \alpha\beta_{ij} = 0$. Suppose that only a random sample of $q = 3$ levels of factor B is included in any single experiment. Suppose that the levels b_2, b_4 , and b_5 are included in an experiment which is actually conducted. For this experiment

$$\overline{\alpha\beta}_i = \frac{\sum_j \alpha\beta_{ij}}{q} = \frac{\alpha\beta_{i2} + \alpha\beta_{i4} + \alpha\beta_{i5}}{3} = \frac{5}{3}.$$

For a large number of random replications of this experiment, where an independent random sample of the levels of factor B is drawn for each replication, the expected value of $\overline{\alpha\beta}_i$ will be zero. The variance of the distribution of $\overline{\alpha\beta}_i$ generated by this kind of sampling procedure depends upon n, q, Q , and the variance $\sigma_{\alpha\beta}^2$.

The notation $\overline{\alpha\beta}_j$ denotes the average of the $\alpha\beta_{ij}$ effects over the levels of factor A present in an experiment at level b_j of factor B . If factor A is a fixed factor, $\overline{\alpha\beta}_j = 0$. However, if factor A is not a fixed factor, $\overline{\alpha\beta}_j$ is not necessarily equal to zero for any single experiment. Over a large number of random replications of the experiment, the expected value of $\overline{\alpha\beta}_j$ will be equal to zero. If factor A is a fixed factor, $\sigma_{\overline{\alpha\beta}_j}^2$ will be zero, since $\overline{\alpha\beta}_j$ will be zero for all replications. If factor A is a random factor, $\sigma_{\overline{\alpha\beta}_j}^2$ will be a function of n, p, P , and $\sigma_{\alpha\beta}^2$.

The mean square due to the main effects of factor A in the experiment is defined to be

$$MS_a = \frac{nq \sum (\bar{A}_i - \bar{G})^2}{p - 1}.$$

An equivalent definition in terms of differences between pairs of means is

$$MS_a = \frac{nq \sum (\bar{A}_i - \bar{A}_{i'})^2}{p(p - 1)}.$$

The summation in this last expression is with respect to all distinct pairs of means, no pair being included twice. The multiplier nq is the number of observations in each \bar{A}_i . The expected value of $\bar{A}_i - \bar{A}_{i'}$ (for independent, random replications of the experiment) is

$$\begin{aligned} E(\bar{A}_i - \bar{A}_{i'}) &= E(\alpha_i - \alpha_{i'}) + E(\overline{\alpha\beta}_i - \overline{\alpha\beta}_{i'}) + E(\bar{\epsilon}_i - \bar{\epsilon}_{i'}) \\ &= \alpha_i - \alpha_{i'} + 0 + 0. \end{aligned}$$

The expected value of MS_a may be shown to be

$$E(MS_a) = \left(\frac{1-n}{N}\right)\sigma_\varepsilon^2 + \left(\frac{1-q}{Q}\right)n\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2.$$

Detailed steps in the derivation of this latter expected value are given by Cornfield and Tukey (1956).

The mean square due to factor B is defined to be

$$\begin{aligned} MS_b &= \frac{np\Sigma(\bar{B}_j - \bar{G})^2}{q-1} \\ &= \frac{np\Sigma(\bar{B}_j - \bar{B}_{j'})^2}{q(q-1)}. \end{aligned}$$

The multiplier np represents the number of observations in each \bar{B}_j . The expected value of this mean square in terms of the parameters in (1) is given in Table 5.6-2.

Table 5.6-2 Expected Values of Mean Squares

Mean square as obtained from experimental data	Expected value of mean square in terms of parameters of (1)
MS_a	$(1 - n/N)\sigma_\varepsilon^2 + n(1 - q/Q)\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2$
MS_b	$(1 - n/N)\sigma_\varepsilon^2 + n(1 - p/P)\sigma_{\alpha\beta}^2 + np\sigma_\beta^2$
MS_{ab}	$(1 - n/N)\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$
MS_{error}	$(1 - n/N)\sigma_\varepsilon^2$

The mean square due to interaction effects in the experiment is defined as

$$MS_{ab} = \frac{n\Sigma\Sigma(\bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G})^2}{(p-1)(q-1)}.$$

An equivalent definition can be given in terms of differential interaction effects. The multiplier n is the number of observations in each \bar{AB}_{ij} . The expected value of this mean square represents the average of the MS_{ab} computed from a large number of independent, random replications of the experiment; this average value is expressed in terms of the parameters in the general linear model (1), which is assumed to represent the sources of variance underlying an observation.

Certain of the coefficients in Table 5.6-2 are either zero or unity, depending upon whether a factor is fixed or random. It will generally be assumed that the number of elements observed in an experiment is small relative to the number of potential elements in the population of elements, i.e., that the ratio n/N for all practical purposes is equal to zero. Hence the coefficient $1 - n/N$ is assumed to be equal to unity.

If factor A is fixed, the ratio p/P will be equal to unity and the coefficient $1 - p/P$ will be equal to zero. If, on the other hand, factor A is random,

the ratio p/P will be equal to zero and the coefficient $1 - p/P$ will be equal to unity. In an analogous manner, the coefficient $1 - q/Q$ is equal to unity when factor B is a fixed factor and equal to zero when factor B is a random factor.

Special cases of the expected values of the mean squares are summarized in Table 5.6-3. Each of these cases is obtained from the general values given in Table 5.6-2 by evaluating the coefficients which depend upon the ratios n/N , p/P , and q/Q . Several different approaches may be used to obtain the special cases given in Table 5.6-3. Specialization of the generalized approach represented by Table 5.6-2 provides the simplest method of attack on the evaluation of more complex experimental designs.

Table 5.6-3 Special Cases of Expected Values of Mean Squares

Mean squares	Case 1 Factor A fixed Factor B fixed	Case 2 Factor A fixed Factor B random	Case 3 Factor A random Factor B random
MS_a	$\sigma_e^2 + nq\sigma_\alpha^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2$
MS_b	$\sigma_e^2 + np\sigma_\beta^2$	$\sigma_e^2 + np\sigma_\beta^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2 + np\sigma_\beta^2$
MS_{ab}	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$
MS_{error}	σ_e^2	σ_e^2	σ_e^2

Case 1, in which both factors are fixed, has been designated by Eisenhart (1947) as model I. Case 2, in which one factor is fixed and the second is random, is called the mixed model. Case 3, in which both factors are random, is called model III, or the *variance components* model. Model I has been more extensively studied than the other two models. In its most general form, the statistical principles underlying model I are identical to those underlying the general regression model having any number of fixed variates and one random variate. As such, the best estimates of various parameters can readily be obtained by the method of least squares. For the case of the generalized model I, application of the method of least squares is straightforward and leads to no difficulties. For the generalized mixed model, application of the principles of maximum likelihood are more direct.

Since the statistical tests made on the experimental data depend upon what these expected values are assumed to be, it is particularly important to specify the conditions under which these expected values are derived. To obtain the general expected values given in Table 5.6-2, the following assumptions are made:

1. There is a population of size P and variance σ_α^2 of main effects of factor A , of which the effects ($\alpha_1, \alpha_2, \dots, \alpha_p$) occurring in the experiment constitute a random sample (sampling without replacement) of size p . The sample may include all the levels of factor A in the population; that is, p may be equal to P .

2. There is a population of size Q and variance σ_β^2 of main effects of factor B , of which the effects $(\beta_1, \beta_2, \dots, \beta_q)$ occurring in the experiment constitute a random sample of size q . The sample may include all the levels of factor B in the population; that is, q may be equal to Q .

3. There is a population of interaction effects of size PQ and variance $\sigma_{\alpha\beta}^2$; the $\alpha\beta_{ij}$'s which occur in the experiment correspond to the combinations of the levels of factor A and factor B that occur in the experiment. That is, one does not have a random sample of pq interaction effects; rather, the interaction effects in the experiment are tied to the levels of factor A and factor B that occur in the experiment. It is assumed that the average (in the population) of the interaction effects over all levels of one factor is independent of the main effects of the other factor; that is, $\bar{\alpha}\beta_i$ is independent of α_i , and $\bar{\alpha}\beta_j$ is independent of β_j .

4. The sampling of the levels of factor A is independent of the sampling of the levels of factor B .

5. The experimental error is independent of all main effects and all interactions. Further, within each cell in the population, ϵ , the experimental error, is assumed to be normally distributed, with mean equal to zero and variance equal to σ_ϵ^2 for all cells in the population.

6. The n observations within each cell of the experiment constitute a random sample of size n from a population of size N (assumed infinite in most cases). The n observations within each cell constitute independent random samples from a random sample of npq independent elements drawn from the basic population.

For purposes of deriving the expected values of mean squares, some of these assumptions may be relaxed. The assumption of normality of the distribution of the experimental error is not required for the derivation of the expected values. However, all these assumptions are needed for the validity of the tests involving the use of F ratios, which are based upon the expected values of the mean squares. In particular, the assumption that the distribution of the experimental error is normal is required in order that the sampling distributions of mean squares be chi-square distributions.

Under the conditions that have been stated, the mean squares computed in the analysis of variance have the following sampling distributions:

<i>Statistic</i>	<i>Sampling distribution</i>
$(p - 1)MS_a/E(MS_a)$	Chi square with $p - 1$ df
$(q - 1)MS_b/E(MS_b)$	Chi square with $q - 1$ df
$(p - 1)(q - 1)MS_{ab}/E(MS_{ab})$	Chi square with $(p - 1)(q - 1)$ df

Principles Underlying Derivation of Expected Values for Mean Squares. To provide some insight into the principles underlying the derivation of the expected values for mean squares, a nonrigorous derivation will be outlined.

From finite sampling theory, one has the following theorem:

$$(1) \quad \sigma_X^2 = \frac{N - n}{N - 1} \frac{\sigma_X^2}{n}.$$

Under random sampling without replacement after each draw, (1) relates the square of the standard error of a mean, σ_X^2 , to the population size N , the sample size n , and the variance σ_X^2 of the variable X in the population.

In (1), the population variance is defined to be

$$\sigma_X^2 = \frac{\Sigma(X - \mu)^2}{N}.$$

If one uses as the definition of the population variance

$$\sigma_X^2 = \frac{\Sigma(X - \mu)^2}{N - 1},$$

then (1) has the form

$$(2) \quad \sigma_X^2 = \left(1 - \frac{n}{N}\right) \frac{\sigma_X^2}{n}.$$

To simplify the notation, the present development will define a population variance using $N - 1$ as the divisor. This definition is consistent with that used by Cornfield and Tukey (1956).

In terms of the right-hand side of the structural model given in (1) of the last section, the mean of all observations made at level a_i in an experiment is

$$(3) \quad \bar{A}_i = \mu_{..} + \alpha_i + \bar{\beta} + \bar{\alpha}\bar{\beta}_i + \bar{\epsilon}_i.$$

In this notation $\bar{\beta}$ represents the average effect of all levels of factor B included in the experiment; $\bar{\beta}$ is constant for all levels of factor A . The notation $\bar{\alpha}\bar{\beta}_i$ represents the average interaction effect associated with level a_i ; $\bar{\alpha}\bar{\beta}_i$ may differ for the various levels of factor A . The notation $\bar{\epsilon}_i$ denotes the average experimental error associated with each \bar{A}_i .

The variance of \bar{A}_i is defined to be

$$s_{\bar{A}}^2 = \frac{\Sigma(\bar{A}_i - \bar{G})^2}{p - 1}.$$

Under the assumption that the terms on the right-hand side of (3) are statistically independent, the expected value of the variance of the left-hand side of (3) will be equal to the sum of the variances of the terms on the right-hand side. Terms which are constants have zero variance. Thus,

$$(4) \quad E(s_{\bar{A}}^2) = \sigma_\alpha^2 + \sigma_{\alpha\beta}^2 + \sigma_\epsilon^2.$$

The mean square due to the main effect of factor A may be written as

$$MS_a = nq s_{\bar{A}}^2.$$

Hence (4) becomes

$$(5) \quad E(MS_a) = E(nqs_A^2) = nqE(s_A^2) = nq\sigma_\alpha^2 + nq\sigma_{\alpha\beta}^2 + nq\sigma_\varepsilon^2.$$

The analogue of the theorem stated in (1) may now be applied to each of the variances in (5) which have a mean as a subscript. Thus, $\bar{\varepsilon}$ is the mean of the experimental error associated with the nq observations from which \bar{A}_i is computed. Therefore,

$$nq\sigma_\varepsilon^2 = \frac{nq(1 - n/N)\sigma_\varepsilon^2}{nq} = \left(1 - \frac{n}{N}\right)\sigma_\varepsilon^2.$$

(The experimental error in the basic linear model is associated with each element in the basic population from which the elements are drawn. Hence the sampling fraction associated with the experimental error is n/N .)

The mean interaction effect in (4) represents the average over the q values of $\alpha\beta_{ij}$ present at level a_i . Hence $\bar{\alpha\beta}_i$ may be considered to be the mean of a sample of size q levels from a population of size Q levels of factor B . The analogue of the theorem in (1) now takes the form

$$nq\sigma_{\alpha\beta}^2 = \frac{nq(1 - q/Q)\sigma_{\alpha\beta}^2}{q} = n\left(1 - \frac{q}{Q}\right)\sigma_{\alpha\beta}^2.$$

Since the main effects and interaction effects are assumed to be independent, restricting the sample of q values of $\alpha\beta_{ij}$ to level a_i does not invalidate the theorem summarized in (1).

The expression for the expected value of the mean square of the main effect of factor A may now be written as

$$(6) \quad E(MS_a) = \left(1 - \frac{n}{N}\right)\sigma_\varepsilon^2 + n\left(1 - \frac{q}{Q}\right)\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2.$$

Each of the variances in (6) is associated with the parameters in the general linear model, whereas in (5) some of the variances were in terms of parameters which are not explicitly in the general linear model. The purpose of defining MS_a as nqs_A^2 now becomes clear. By adding the multiplier nq , one may conveniently express each of the variances in (5) in terms of parameters which are explicit in the linear model.

Derivation of the expected values of the mean squares for the main effects of B and for the AB interaction follows the same general line of reasoning. The algebraic procedures whereby the various means are obtained from the experimental data are carried through in terms of the right-hand side of the basic structural model for an observation. Then, using the assumptions underlying the model, one obtains the expected values of the mean squares by the principles that have just been outlined.

The principles underlying the derivation of expected values of the mean squares will be illustrated for the case of a 2×2 factorial experiment having n observations in each cell. In terms of a general linear model, the cell

means and the marginal means for the levels of factor B may be represented as follows:

$$\begin{array}{rcl} & b_1 & b_2 \\ \overline{AB}_{11} &= \mu_{..} + \alpha_1 + \beta_1 + \alpha\beta_{11} + \bar{\epsilon}_{11} & \overline{AB}_{12} = \mu_{..} + \alpha_1 + \beta_2 + \alpha\beta_{12} + \bar{\epsilon}_{12} \\ \overline{AB}_{21} &= \mu_{..} + \alpha_2 + \beta_1 + \alpha\beta_{21} + \bar{\epsilon}_{21} & \overline{AB}_{22} = \mu_{..} + \alpha_2 + \beta_2 + \alpha\beta_{22} + \bar{\epsilon}_{22} \\ \bar{B}_1 &= \mu_{..} + \bar{\alpha} + \beta_1 + \bar{\alpha}\beta_{.1} + \bar{\epsilon}_{.1} & \bar{B}_2 = \mu_{..} + \bar{\alpha} + \beta_2 + \bar{\alpha}\beta_{.2} + \bar{\epsilon}_{.2} \end{array}$$

The difference between the two marginal means for factor B estimates the following parameters:

$$\bar{B}_1 - \bar{B}_2 = (\beta_1 - \beta_2) + (\bar{\alpha}\beta_{.1} - \bar{\alpha}\beta_{.2}) - (\bar{\epsilon}_{.1} - \bar{\epsilon}_{.2}).$$

Multiplying each side of the above expression by np gives

$$np(\bar{B}_1 - \bar{B}_2) = np(\beta_1 - \beta_2) + np(\bar{\alpha}\beta_{.1} - \bar{\alpha}\beta_{.2}) + np(\bar{\epsilon}_{.1} - \bar{\epsilon}_{.2}).$$

Since the terms on the right-hand side of the above expression are assumed to be statistically independent, the variance of the term on the left-hand side will estimate the sum of the variances of the terms on the right-hand side. Hence,

$$E(\text{MS}_b) = np\sigma_\beta^2 + np\sigma_{\alpha\beta}^2 + np\sigma_{\bar{\epsilon}}^2.$$

Using the analogue of the relation given in (1),

$$np\sigma_{\alpha\beta}^2 = n\left(1 - \frac{p}{P}\right)\sigma_{\alpha\beta}^2, \quad \text{and} \quad np\sigma_{\bar{\epsilon}}^2 = \left(1 - \frac{n}{N}\right)\sigma_{\bar{\epsilon}}^2.$$

Thus,
$$E(\text{MS}_b) = np\sigma_\beta^2 + n\left(1 - \frac{p}{P}\right)\sigma_{\alpha\beta}^2 + \left(1 - \frac{n}{N}\right)\sigma_{\bar{\epsilon}}^2.$$

In most experimental designs the potential number of experimental units, N , is considered to be infinite. Hence $1 - n/N = 1$.

In a $p \times q$ factorial experiment having n observations per cell, least-squares estimates of the terms in the general linear model are obtained in a relatively simple manner, provided that all factors are fixed. Let m , a_i , b_j , and ab_{ij} be estimates of the parameters μ , α_i , β_j , and $\alpha\beta_{ij}$. The least-squares estimates are obtained from the condition

$$(7) \quad \sum_i \sum_j \sum_k (X_{ijk} - a_i - b_j - ab_{ij} - m)^2 = \text{minimum},$$

subject to the restrictions that

$$\sum a_i = 0, \quad \sum b_j = 0, \quad \sum_i ab_{ij} = 0, \quad \sum_j ab_{ij} = 0.$$

Corresponding restrictions on the parameters follow directly from the definitions under model I.

The solutions to the normal equations obtained from (7) are

$$\begin{aligned} m &= \bar{G}, \\ a_i &= \bar{A}_i - \bar{G}, \\ b_j &= \bar{B}_j - \bar{G}, \\ ab_{ij} &= \bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G}. \end{aligned}$$

These estimators are identical to those obtained by working directly from the linear model. When the cell frequencies are not equal, the least-squares estimators do not have this simple form. The latter case is considered in Sec. 5.25.

5.7 Principles for Constructing F Ratios

For either model I (all factors assumed fixed) or model III (all factors assumed random) the sampling distributions of the mean squares for main effects and interactions can be adequately approximated by independent chi-square distributions. For model II (some factors fixed, some factors random) all required sampling distributions are independent chi-square distributions only if highly restrictive assumptions on covariances are made. The principles for testing hypotheses to be presented in this section hold rigorously for models I and III. In practice, tests under model II follow the principles to be presented here; however, interpretations under model II require special care. Scheffé (1959, pp. 264, 288), in particular, has questioned principles for constructing F tests of the type to be presented here for special cases of model II.

The hypothesis that $\alpha_1 = \alpha_2 = \cdots = \alpha_P$ (that is, the hypothesis of no differences between the main effects of factor A) is equivalent to the hypothesis that $\sigma_\alpha^2 = 0$. This hypothesis is in turn equivalent to the following hypotheses:

1. All possible comparisons (or contrasts) among the main effects of factor A are equal to zero.

2.

$$\mu_{1.} = \mu_{2.} = \cdots = \mu_{P.} = \mu_{..}$$

To test this hypothesis against the alternative hypothesis that $\sigma_\alpha^2 > 0$ requires the construction of an F ratio. In terms of the expected value of mean squares, the F ratio for this test has the general form

$$E(F) = \frac{u + c\sigma_\alpha^2}{u},$$

where u is some linear function of the variances of other parameters in the model and c is some coefficient. In words, $E(\text{MS}_{\text{numerator}})$ must be equal to $E(\text{MS}_{\text{denominator}})$ when $\sigma_\alpha^2 = 0$.

For the test under consideration, the mean square in the numerator of the F ratio must be MS_a . The mean square that is in the denominator depends upon the expected value of MS_a under the proper model. For model I, the appropriate denominator for this F ratio is MS_{error} ; for model III, the appropriate denominator is MS_{ab} . Thus, in order to form an F ratio in the analysis of variance, knowledge of expected value of mean squares under the appropriate model is needed. This, in essence, implies that the F ratio depends upon the design of the experiment.

If the numerator and denominator of an F ratio satisfy the structural requirements in terms of the expected values of the mean squares, and if the sampling distributions of these mean squares are independent chi squares when the hypothesis being tested is true, then the resulting F ratio will have a sampling distribution approximated by an F distribution. The degrees of freedom for the resulting F distribution are, respectively, the degrees of freedom for the numerator mean square and the degrees of freedom for the denominator mean square. General principles for setting up F ratios are illustrated in Table 5.7-1.

Table 5.7-1 Tests of Hypotheses under Model III

Source of variation	E(MS)	Hypothesis being tested	F ratio
Main effect of factor A	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2$	$H_1: \sigma_\alpha^2 = 0$	$F = MS_a/MS_{ab}$
Main effect of factor B	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + np\sigma_\beta^2$	$H_1: \sigma_\beta^2 = 0$	$F = MS_b/MS_{ab}$
$A \times B$ interaction	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2$	$H_1: \sigma_{\alpha\beta}^2 = 0$	$F = MS_{ab}/MS_{\text{error}}$
Error	σ_ϵ^2		

The expected values in this table are those appropriate for model III. In terms of these expected values the F ratio used to test the hypothesis that $\sigma_\alpha^2 = 0$ has the form

$$E(F) = \frac{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2 + nq\sigma_\alpha^2}{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2}.$$

When the hypothesis being tested is true (when $\sigma_\alpha^2 = 0$), numerator and denominator of the expected value of the F ratio are identical. Thus, $E(MS_a) = E(MS_{ab})$ when $\sigma_\alpha^2 = 0$. When σ_α^2 is greater than zero, the expected value of the F ratio is greater than unity by an amount which depends upon the term $nq\sigma_\alpha^2$. From the structure of the F ratio, it can be less than unity only because of sampling error associated with MS_a and MS_{ab} ; for any single experiment each of these statistics may be independently either less than its expected value or greater than its expected value. Alternatively, the F ratio might be less than unity when some of the assumptions about the model do not hold.

The F ratio appropriate for testing the hypothesis $\sigma_{\alpha\beta}^2 = 0$ has the structure

$$E(F) = \frac{\sigma_\epsilon^2 + n\sigma_{\alpha\beta}^2}{\sigma_\epsilon^2}.$$

When the hypothesis being tested is true, numerator and denominator have the same expected value. When $\sigma_{\alpha\beta}^2 > 0$, the expected value of this F ratio

will be greater than unity. The F ratio as obtained from the experimental data has the form

$$F = \frac{MS_{ab}}{MS_{\text{error}}}.$$

Under the hypothesis that $\sigma_{\alpha\beta}^2 = 0$, the expected value of this F ratio is unity. When $\sigma_{\alpha\beta}^2$ is greater than zero, this F statistic has an expected value greater than unity by an amount which is a function of $n\sigma_{\alpha\beta}^2$. Under the hypothesis that $\sigma_{\alpha\beta}^2 = 0$, the sampling distribution of the F ratio is the F distribution having $(p - 1)(q - 1)$ and $pq(n - 1)$ degrees of freedom.

To evaluate the power of a test using an F ratio requires a knowledge of the sampling distribution of the F ratio for specified nonzero values of the variance in the hypothesis being tested. For nonzero values of this variance, the sampling distribution of the F ratio is not the ordinary F distribution but rather the *noncentral* F distribution. Convenient charts are available (Table B.11) to determine the sample size required for a predetermined power with respect to specified alternative hypotheses. A collection of charts for determining the power of F tests also appears in Scheffé (1959, pp. 438-455).

5.8 Higher-order Factorial Experiments

When a factorial experiment includes three or more factors, different orders of interaction are possible. For example, in a $2 \times 3 \times 5$ factorial experiment, having 10 independent observations in each cell, the analysis of variance generally has the form given in Table 5.8-1.

Table 5.8-1 Analysis of Variance for $2 \times 3 \times 5$ Factorial Experiment Having 10 Observations per Cell

Source of variation	Sum of squares	df	df (general)
<i>A</i> main effects	SS_a	1	$p - 1$
<i>B</i> main effects	SS_b	2	$q - 1$
<i>C</i> main effects	SS_c	4	$r - 1$
<i>AB</i> interaction	SS_{ab}	2	$(p - 1)(q - 1)$
<i>AC</i> interaction	SS_{ac}	4	$(p - 1)(r - 1)$
<i>BC</i> interaction	SS_{bc}	8	$(q - 1)(r - 1)$
<i>ABC</i> interaction	SS_{abc}	8	$(p - 1)(q - 1)(r - 1)$
Experimental error (within cell)	SS_{error}	270	$pqr(n - 1)$
Total	SS_{total}	299	$npqr - 1$

In a three-factor experiment there are three interactions which involve two factors: $A \times B$, $A \times C$, $B \times C$. There is one three-factor interaction. The $A \times B \times C$ interaction represents the unique effects attributable to the combination of the three factors, i.e., the effects that cannot be predicted

from a knowledge of the main effects and all two-factor interactions. The notation that was introduced for the case of a two-factor experiment can be extended as follows:

	Levels in population	Levels in experiment
Factor <i>A</i>	$a_1, a_2, \dots, a_I, \dots, a_P$	$a_1, a_2, \dots, a_i, \dots, a_p$
Factor <i>B</i>	$b_1, b_2, \dots, b_J, \dots, b_Q$	$b_1, b_2, \dots, b_j, \dots, b_q$
Factor <i>C</i>	$c_1, c_2, \dots, c_K, \dots, c_R$	$c_1, c_2, \dots, c_k, \dots, c_r$

The definitions of fixed and random factor given in Sec. 5.2 also apply to factor *C*. If $r = R$, then factor *C* is a fixed factor. If the r levels of factor *C* in the experiment are a random sample from the R levels in the population, and if R is quite large relative to r , then factor *C* is a random factor. If the R levels are reduced to $R_{\text{effective}}$ levels by some systematic, nonrandom procedure, then factor *C* is considered fixed when $r = R_{\text{effective}}$.

The notation for cell means used for the case of a two-factor experiment may also be extended. An observation on element m under treatment combination abc_{ijk} is designated by X_{ijkm} . Notation for cell means is summarized in Table 5.8-2. In this notation system, μ_{ijk} designates the mean

Table 5.8-2 Notation for Means in a Three-factor Experiment

	Population mean	Experiment mean
Elements in cell abc_{ijk}	μ_{ijk}	\overline{ABC}_{ijk}
Elements under ab_{ij}	$\mu_{ij.}$	\overline{AB}_{ij}
Elements under ac_{ik}	$\mu_{i..k}$	\overline{AC}_{ik}
Elements under bc_{jk}	$\mu_{.jk}$	\overline{BC}_{jk}
Elements under a_i	$\mu_{i..}$	\bar{A}_i
Elements under b_j	$\mu_{.j.}$	\bar{B}_j
Elements under c_k	$\mu_{..k}$	\bar{C}_k

of the N potential observations that could be made under treatment combination abc_{ijk} . The notation $\mu_{ij.}$ designates the mean of the NR potential observations that could be made under the treatment combinations $abc_{ij1}, abc_{ij2}, \dots, abc_{ijR}$ (N potential observations under each of the R treatment combinations). In terms of symbols,

$$\mu_{ij.} = \frac{\sum_K \mu_{ijK}}{R}$$

(The subscript K is used here to indicate that the average is over all potential levels of factor *C* and not just those in any single experiment.)

The notation $\mu_{i..}$ designates the mean of the NQR potential observations that could be made under the treatment combinations in which factor A is at level a_i . Thus,

$$\mu_{i..} = \frac{\sum_J \sum_K \mu_{iJK}}{QR}.$$

(The subscripts J and K indicate that the average is over all levels of factors B and C , not just those included in the experiment.) Similarly, $\mu_{..k}$ denotes the mean of the potential NPQ observations that could be made under level c_k , that is,

$$\mu_{..k} = \frac{\sum_I \sum_J \mu_{IJK}}{PQ}.$$

The notation k refers to a level of factor C actually included in an experiment. If all the factors in an experiment are fixed factors, there is no need to make the distinction between I, J , and K and i, j , and k .

The numerical data given in Table 5.8-3 will be used to illustrate the definitions of main effects and interactions in a three-factor experiment. The data in this table include all levels of each of the factors, and the entries in the cells are the means of all the potential observations that could be in each of the cells. Thus the numerical entries represent the parameters for a specified population.

In part i the cell entries may be designated by the symbol μ_{IJK} . For example, $\mu_{211} = 60$. In symbols,

$$\mu_{211} = \frac{\sum_M X_{211M}}{N} = 60.$$

The entries along the lower margin of part i represent the means of the respective columns. Thus the entry 40 at the extreme left represents the mean of all potential observations in which factor B is at level b_1 and factor C is at level c_1 ; in symbols, this mean is $\mu_{.11}$. Thus,

$$\mu_{.11} = \frac{\sum_I \mu_{I11}}{P} = \frac{20 + 60}{2} = 40.$$

Part ia summarizes the symbols for the entries in part i.

In part ii each of the numerical entries in the cells represents a mean which has the general symbol $\mu_{I..K}$. For example, the entry in cell ac_{21} is

$$\mu_{2..1} = \frac{\sum_J \mu_{2J1}}{Q} = \frac{60 + 40 + 50}{3} = 50.$$

Each of the marginal entries to the right of the cells in part ii has the general symbol $\mu_{I..}$. For example, the entry 30 is the mean of the entries in row a_1 ; in symbols, $\mu_{1..} = 30$. Thus,

$$\mu_{1..} = \frac{\sum_K \mu_{1.K}}{R} = \frac{40 + 20}{2} = 30.$$

The entries along the bottom margin of part ii represent the means of all potential observations at specified levels of factor C. Thus,

$$\mu_{..2} = \frac{\sum_I \mu_{I.2}}{P} = \frac{20 + 30}{2} = 25.$$

Part iia summarizes the symbols for corresponding entries in part ii.

In part iii each of the cell entries has the general designation $\mu_{IJ.}$. The marginal entries at the right may be designated $\mu_{I..}$; the marginal entries

Table 5.8-3 Population Means for $2 \times 3 \times 2$ Factorial Experiment

(i)		b_1		b_2		b_3	
		c_1	c_2	c_1	c_2	c_1	c_2
	a_1	20	0	30	10	70	50
	a_2	60	40	40	20	50	30
		40	20	35	15	60	40

(ii)		c_1	c_2	Mean	(iii)		b_1	b_2	b_3	Mean
	a_1	40	20	30		a_1	10	20	60	30
	a_2	50	30	40		a_2	50	30	40	40
		45	25	35			30	25	50	35

(ia)		b_1		b_2		b_3		
		c_1	c_2	c_1	c_2	c_1	c_2	Mean
	a_1	μ_{111}	μ_{112}	μ_{121}	μ_{122}	μ_{131}	μ_{132}	$\mu_{1..}$
	a_2	μ_{211}	μ_{212}	μ_{221}	μ_{222}	μ_{231}	μ_{232}	$\mu_{2..}$
		$\mu_{..1}$	$\mu_{..2}$	$\mu_{.21}$	$\mu_{.22}$	$\mu_{.31}$	$\mu_{.32}$	

(iia)		c_1	c_2	Mean	(iiia)		b_1	b_2	b_3	Mean
	a_1	$\mu_{1.1}$	$\mu_{1.2}$	$\mu_{1..}$		a_1	$\mu_{11.}$	$\mu_{12.}$	$\mu_{13.}$	$\mu_{1..}$
	a_2	$\mu_{2.1}$	$\mu_{2.2}$	$\mu_{2..}$		a_2	$\mu_{21.}$	$\mu_{22.}$	$\mu_{23.}$	$\mu_{2..}$
		$\mu_{..1}$	$\mu_{..2}$	$\mu_{...}$			$\mu_{.1.}$	$\mu_{.2.}$	$\mu_{.3.}$	$\mu_{...}$

at the bottom may be designated $\mu_{..j}$. Thus the entry 50 at the bottom of part iii is

$$\mu_{..3} = \frac{\sum \mu_{I3.}}{P} = \frac{60 + 40}{2} = 50.$$

The main effects and interaction effects for a three-factor experiment will be defined in terms of the data in Table 5.8-3. The main effect due to level a_1 is

$$\alpha_1 = \mu_{1..} - \mu_{...} = 30 - 35 = -5.$$

In words, the main effect of level a_1 is a measure of the extent to which the mean of all potential observations at level a_1 , averaged over all potential levels of factors B and C , differs from the grand mean of all potential observations. In general the main effect of level a_i is

$$\alpha_i = \mu_{i..} - \mu_{...}$$

The main effect of level b_3 of factor B is by definition

$$\beta_3 = \mu_{..3} - \mu_{...} = 50 - 35 = 15.$$

In general the main effect of level b_j is

$$\beta_j = \mu_{..j} - \mu_{...}$$

The main effect for level c_k of factor C has the general definition

$$\gamma_k = \mu_{..k} - \mu_{...}$$

The interaction effect of level a_i with level b_j is the definition

$$\begin{aligned} \alpha\beta_{ij} &= \mu_{ij.} - \mu_{...} - \alpha_i - \beta_j = \mu_{ij.} - (\mu_{...} + \alpha_i + \beta_j) \\ &= \mu_{ij.} - \mu_{i..} - \mu_{..j} + \mu_{...} \end{aligned}$$

In words, $\alpha\beta_{ij}$ measures the extent to which the mean of all potential observations differs from the sum of the main effects of a_i and b_j and the grand mean. For example,

$$\alpha\beta_{13} = 60 - 30 - 50 + 35 = 15.$$

The interaction effect of level a_i with c_k is

$$\begin{aligned} \alpha\gamma_{ik} &= \mu_{i.k} - \mu_{...} - \alpha_i - \gamma_k \\ &= \mu_{i.k} - \mu_{i..} - \mu_{..k} + \mu_{...} \end{aligned}$$

The interaction effect of level b_j with c_k is

$$\begin{aligned} \beta\gamma_{jk} &= \mu_{.jk} - \mu_{...} - \beta_j - \gamma_k \\ &= \mu_{.jk} - \mu_{.j.} - \mu_{..k} + \mu_{...} \end{aligned}$$

For example, the interaction of level b_3 with level c_1 is

$$\beta\gamma_{31} = 60 - 50 - 45 + 35 = 0.$$

The interaction of a_i , b_j , and c_k is defined to be

$$\begin{aligned}\alpha\beta\gamma_{ijk} &= \mu_{ijk} - \mu_{...} - (\alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha_i + \beta_j + \gamma_k) \\ &= \mu_{ijk} - \mu_{ij.} - \mu_{i.k} - \mu_{.jk} + \mu_{i..} + \mu_{.j.} + \mu_{..k} - \mu_{...}\end{aligned}$$

In words, the interaction of three factors measures the difference between the mean of all potential observations under a specified combination, μ_{ijk} , and the sum of two-factor interactions, main effects, and the grand mean. The three-factor interaction is in essence a measure of the nonadditivity of two-factor interactions and main effects. For example, for the data in Table 5.8-3, the interaction of a_1 , b_3 , and c_2 is

$$\begin{aligned}\alpha\beta\gamma_{132} &= \mu_{132} - \mu_{13.} - \mu_{1.2} - \mu_{.32} + \mu_{1..} + \mu_{.3.} + \mu_{..2} - \mu_{...} \\ &= 50 - 60 - 20 - 40 + 30 + 50 + 25 - 35 \\ &= 0.\end{aligned}$$

$$\begin{aligned}\text{Equivalently, } \alpha\beta\gamma_{132} &= \mu_{132} - \mu_{...} - (\alpha\beta_{13} + \alpha_{12} + \beta_{32} + \alpha_1 + \beta_3 + \gamma_2) \\ &= 50 - 35 - [15 + 0 + 0 + (-5) + 15 + (-10)] \\ &= 0.\end{aligned}$$

From the definitions of these effects it follows that

$$\begin{aligned}\sum_I \alpha_I &= 0, \quad \sum_J \beta_J = 0, \quad \sum_K \gamma_K = 0; \\ \sum_I \alpha\beta_{IJ} &= \sum_J \alpha\beta_{IJ} = \sum_I \alpha\gamma_{IK} = \sum_K \alpha\gamma_{IK} = \sum_J \beta\gamma_{JK} = \sum_K \beta\gamma_{JK} = 0; \\ \sum_I \alpha\beta\gamma_{IJK} &= \sum_J \alpha\beta\gamma_{IJK} = \sum_K \alpha\beta\gamma_{IJK} = 0.\end{aligned}$$

In each case the summation is over all the potential levels of the factors. If, for example, $k \neq K$, then

$$\sum_k \gamma_k \neq 0.$$

The variance due to the main effects of factor A is

$$\sigma_\alpha^2 = \frac{\sum_I \alpha_I^2}{P - 1}.$$

Similarly, the variances due to the main effects of factors B and C are, respectively,

$$\begin{aligned}\sigma_\beta^2 &= \frac{\sum_J \beta_J^2}{Q - 1}, \\ \sigma_\gamma^2 &= \frac{\sum_K \gamma_K^2}{R - 1}.\end{aligned}$$

When the main effects within any factor are all equal, the variance corresponding to these main effects will be zero. Hence the equality of main effects implies that the variance corresponding to these main effects is zero.

The variance due to two-factor interactions is defined as follows:

$$\sigma_{\alpha\beta}^2 = \frac{\sum_I \sum_J (\alpha\beta_{IJ})^2}{(P-1)(Q-1)},$$

$$\sigma_{\alpha\gamma}^2 = \frac{\sum_I \sum_K (\alpha\gamma_{IK})^2}{(P-1)(R-1)},$$

$$\sigma_{\beta\gamma}^2 = \frac{\sum_J \sum_K (\beta\gamma_{JK})^2}{(Q-1)(R-1)}.$$

The variance due to the three-factor interaction is

$$\sigma_{\alpha\beta\gamma}^2 = \frac{\sum_I \sum_J \sum_K (\alpha\beta\gamma_{IJK})^2}{(P-1)(Q-1)(R-1)}.$$

The variance due to the experimental error has the same general definition as that given for two-factor experiments; it is the variance of the measurements on the N potential elements within each potential cell of the experiment. Thus, for cell abc_{ijk} ,

$$\sigma_{\varepsilon_{ijk}}^2 = \frac{\sum_M (X_{ijkM} - \mu_{ijk})^2}{N-1},$$

where the subscript M represents a potential element in the cell specified. Assuming that the variance due to experimental error is equal for all potential cells in the experiment, the over-all variance due to experimental error is

$$\sigma_{\varepsilon}^2 = \frac{\sum_I \sum_J \sum_K \sigma_{\varepsilon_{IJK}}^2}{PQR}.$$

Mean squares as computed from the actual experimental data are estimates of linear functions of these population variances. The mean squares of main effects, as obtained from observed data, will estimate a linear function of the variance due to the specified main effects, variance due to interactions effects, and variance due to experimental error.

The extension of the notation and definitions to four-factor experiments is direct. For example, the main effect of level d_m of factor D is

$$\delta_m = \mu_{...m} - \mu_{....}$$

This main effect is estimated by

$$\text{est}(\delta_m) = \bar{D}_m - \bar{G}.$$

The notation $\text{est}(\delta_m)$ denotes an estimate of the parameter δ_m obtained from the data in the experiment. The interaction effect associated with levels a_i and d_m is

$$\alpha\delta_{im} = \mu_{i..m} - \mu_{....} - (\alpha_i + \delta_m).$$

This interaction effect is estimated by

$$\text{est } (\alpha\delta_{im}) = \overline{AD}_{im} - \bar{A}_i - \bar{D}_m + \bar{G}.$$

The interaction of levels a_i , b_j , and d_m is defined to be

$$\alpha\beta\delta_{ijm} = \mu_{ijm} - \mu_{...} - (\alpha\beta_{ij} + \alpha\delta_{im} + \beta\delta_{jm}) + (\alpha_i + \beta_j + \delta_m).$$

This interaction effect is estimated by

$$\text{est } (\alpha\beta\delta_{ijm}) = \overline{ABD}_{ijm} - \overline{AB}_{ij} - \overline{AD}_{im} - \overline{BD}_{jm} + \bar{A}_i + \bar{B}_j + \bar{D}_m - \bar{G}.$$

This last estimate has the general form

$$(\text{3-factor mean}) - \Sigma(\text{2-factor means}) + \Sigma(\text{1-factor means}) - \bar{G}.$$

The estimate of a four-factor interaction has the general form

$$\begin{aligned} &(\text{4-factor mean}) - \Sigma(\text{3-factor means}) + \Sigma(\text{2-factor means}) \\ &\quad - \Sigma(\text{1-factor means}) + \bar{G}. \end{aligned}$$

For example, the interaction of levels a_1, b_2, c_3, d_4 is estimated by

$$\begin{aligned} \text{est } (\alpha\beta\gamma\delta_{1234}) = & \overline{ABCD}_{1234} - (\overline{ABC}_{123} + \overline{ABD}_{124} + \overline{ACD}_{134} + \overline{BCD}_{234}) \\ & + (\overline{AB}_{12} + \overline{AC}_{13} + \overline{AD}_{14} + \overline{BC}_{23} + \overline{BD}_{24} + \overline{CD}_{34}) \\ & - (\bar{A}_1 + \bar{B}_2 + \bar{C}_3 + \bar{D}_4) + \bar{G}. \end{aligned}$$

The term $\Sigma(\text{3-factor means})$ in the general expression for the estimate of a four-factor interaction effect includes all possible means of the form \overline{UVW}_{rst} , where r, s , and t are the subscripts for corresponding terms in the interaction effect being estimated. In a four-factor experiment the number of terms in $\Sigma(\text{3-factor means})$ is equal to the number of combinations of four things taken three at a time, which is symbolized ${}_4C_3$. This number is

$${}_4C_3 = \frac{4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3} = 4.$$

For a four-factor experiment the number of terms in the summation $\Sigma(\text{2-factor means})$ is

$${}_4C_2 = \frac{4 \cdot 3}{1 \cdot 2} = 6.$$

In a k -factor experiment, the estimate of the k -factor interaction effect has the general form

$$\begin{aligned} &(\textit{k-factor mean}) - \Sigma[(\textit{k} - 1)\text{-factor means}] + \Sigma[(\textit{k} - 2)\text{-factor means}] \\ &\quad - \Sigma[(\textit{k} - 3)\text{-factor means}] + \Sigma[(\textit{k} - 4)\text{-factor means}] \\ &\quad - \dots \end{aligned}$$

The last term is $\pm[(\textit{k} - \textit{k})\text{-factor means}] = \pm\bar{G}$. If k is an even number, the last term is $+\bar{G}$; if k is an odd number, the last term is $-\bar{G}$. The number of terms in the summation $\Sigma[(\textit{k} - 1)\text{-factor means}]$ is ${}_kC_{k-1}$; the number of

terms in the summation $\Sigma[(k-2)\text{-factor means}]$ is ${}_k C_{k-2}$. For example, if $k=5$,

$${}_k C_{k-2} = {}_5 C_3 = \frac{5 \cdot 4 \cdot 3}{1 \cdot 2 \cdot 3} = 10; \quad {}_k C_{k-1} = {}_5 C_4 = \frac{5 \cdot 4 \cdot 3 \cdot 2}{1 \cdot 2 \cdot 3 \cdot 4} = 5.$$

5.9 Estimation and Tests of Significance for Three-factor Experiments

For purposes of demonstrating the principles underlying the analysis that will be made, it is convenient to formulate a structural model for an observation. For a three-factor experiment, the structural model has the form

$$(1) \quad X_{ijkm} = f(abc_{ijk}) + \varepsilon_{ijkm}.$$

The observation on element m under treatment combination abc_{ijk} is designated by the symbol X_{ijkm} . The symbol $f(abc_{ijk})$ denotes the hypothetically true effect of the treatment combination abc_{ijk} as measured by population main effects and population interaction effects. The symbol ε_{ijkm} is the experimental error associated with the measurement on element m . In this context the experimental error is considered to be the difference between the observed measurement and the "true" measurement given by $f(abc_{ijk})$.

For purposes of the analysis that follows it will be assumed that the true measurement is a linear function of the main effects and interaction effects. Specifically,

$$(2) \quad f(abc_{ijk}) = \mu_{...} + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk}.$$

Thus, (2) expresses the true measurement as a linear function of main effects and interaction effects associated with a specific treatment combination. If (1) and (2) are combined, the resulting structural model is a generalization of the structural model given in Sec. 5.6 for a two-factor experiment. The assumptions that will be made about this model in the course of the analysis are direct generalizations of those summarized in Sec. 5.6.

It will be assumed that p levels of factor A are selected at random from a population of P levels. It will further be assumed that an independent random sample of q levels of factor B is selected from a population of Q levels and that a third independent random sample of size r is selected from a population of R levels of factor C . The treatments in the experiment are the pqr combinations that result when each of the selected levels of one factor is combined with each of the selected levels from the other factors. For example, if level a_1 is used in combination with b_2 and c_3 , the resulting treatment is designated abc_{123} . In a $p \times q \times r$ factorial experiment there are pqr treatment combinations.

It will also be assumed that a random sample of $npqr$ elements is drawn from a specified population. Random subsamples of size n each are assigned to each of the pqr treatment combinations to be studied in the experiment. After administration of the treatments, each of the elements

is measured on a criterion of effectiveness (the dependent variable). The scale of measurement for the criterion is assumed to be given in terms of an experimentally meaningful unit.

From the data obtained in the experiment, mean squares are computed to estimate variances due to the structural variables on the right-hand side of the structural model. For a three-factor experiment, the definitions of the mean squares for main effects, interactions, and experimental error are summarized in Table 5.9-1. With the exception of multipliers and the ranges of

Table 5.9-1 Definition of Mean Squares

<i>A</i> main effect	$MS_a = nqr \sum_i (\bar{A}_i - \bar{G})^2 / (p - 1)$
<i>B</i> main effect	$MS_b = npr \sum_j (\bar{B}_j - \bar{G})^2 / (q - 1)$
<i>C</i> main effect	$MS_c = npq \sum_k (\bar{C}_k - \bar{G})^2 / (r - 1)$
<i>AB</i> interaction	$MS_{ab} = nr \sum_i \sum_j (\bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G})^2 / (p - 1)(q - 1)$
<i>AC</i> interaction	$MS_{ac} = nq \sum_i \sum_k (\bar{AC}_{ik} - \bar{A}_i - \bar{C}_k + \bar{G})^2 / (p - 1)(r - 1)$
<i>BC</i> interaction	$MS_{bc} = np \sum_j \sum_k (\bar{BC}_{jk} - \bar{B}_j - \bar{C}_k + \bar{G})^2 / (q - 1)(r - 1)$
<i>ABC</i> interaction	$MS_{abc} = n \sum_i \sum_j \sum_k (\bar{ABC}_{ijk} - \bar{AB}_{ij} - \bar{AC}_{ik} - \bar{BC}_{jk} + \bar{A}_i + \bar{B}_j + \bar{C}_k - \bar{G})^2 / (p - 1)(q - 1)(r - 1)$
Experimental error	$MS_{error} = \sum_i \sum_j \sum_k \sum_m (X_{ijkm} - \bar{ABC}_{ijk})^2 / pqr(n - 1)$

summation, these definitions carry out on the experimental data the same operations that would be carried out on the population to obtain the variance due to main effects, interactions, and experimental error.

For example, in the population the variance due to the main effects of factor *A* is

$$\sigma_a^2 = \frac{\sum_I (\mu_{I..} - \mu_{...})^2}{P - 1}.$$

The mean square due to the main effects of factor *A* as computed from the data in the experiment is

$$MS_a = \frac{nqr \sum_i (\bar{A}_i - \bar{G})^2}{p - 1}.$$

As another example, in the population the variance due to the *BC* interaction is

$$\sigma_{\beta_7}^2 = \frac{\sum_J \sum_K (\mu_{.JK} - \mu_{.J.} - \mu_{..K} + \mu_{...})^2}{(Q - 1)(R - 1)}.$$

Table 5.9-2 Expected Value for Mean Squares in a Three-factor Experiment Having n Observations per Cell

		Special Cases ($n/N = 0$ in all cases)		
Mean square		Model I All factors fixed	Model II Factor A random, all others fixed	Model III All factors random
		MS_a $\sigma_e^2 + nqrc_\alpha^2$ MS_b $\sigma_e^2 + nprc_\beta^2$ MS_c $\sigma_e^2 + npqc_\gamma^2$ MS_{ab} $\sigma_e^2 + nr\sigma_{\alpha\beta}^2$ MS_{ac} $\sigma_e^2 + nq\sigma_{\alpha\gamma}^2$ MS_{bc} $\sigma_e^2 + np\sigma_{\beta\gamma}^2$ MS_{abc} $\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$ MS_{error} $(1 - n/N)\sigma_e^2$	$\sigma_e^2 + nqrc_\alpha^2$ $\sigma_e^2 + nprc_\beta^2$ $\sigma_e^2 + npqc_\gamma^2$ $\sigma_e^2 + nr\sigma_{\alpha\beta}^2$ $\sigma_e^2 + nq\sigma_{\alpha\gamma}^2$ $\sigma_e^2 + np\sigma_{\beta\gamma}^2$ $\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$	$\sigma_e^2 + nqrc_\alpha^2$ $\sigma_e^2 + nprc_\beta^2$ $\sigma_e^2 + npqc_\gamma^2$ $\sigma_e^2 + nr\sigma_{\alpha\beta}^2$ $\sigma_e^2 + nq\sigma_{\alpha\gamma}^2$ $\sigma_e^2 + np\sigma_{\beta\gamma}^2$ $\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$

(i)

(ii)

The mean square due to the BC interaction as obtained from the data in the experiment is

$$MS_{bc} = \frac{np \sum_j \sum_k (\overline{BC}_{jk} - \bar{B}_j - \bar{C}_k + \bar{G})^2}{(q-1)(r-1)}.$$

In each case, with the exceptions noted, the mean square duplicates for the data obtained from the experiment the definitions for the variance due to main effects and interactions in the population.

The expected values of the mean squares computed from the experimental data are not, however, equal to a simple multiple of the corresponding variance in the population. For example,

$$E(MS_a) \neq nq\sigma_a^2.$$

In general, the expected value of the mean square due to the main effects of factor A will depend upon variance due to interactions with the levels of factor A , variance due to experimental error, and the variance due to the main effects of factor A . Similarly,

$$E(MS_{bc}) \neq np\sigma_{bc}^2.$$

In general the expected value of MS_{bc} will depend upon the variance due to higher-order interactions which involve factors B and C , variance due to experimental error, and the variance due to the BC interaction. The expected values for the mean squares defined in Table 5.9-1 are summarized in Table 5.9-2. Part i of this table gives general values; part ii specializes these general values for specific cases. Case 2 is one of several possible special cases of the mixed model; for example, two factors may be random and one factor fixed; or one factor may be random, and two factors may be fixed. In all these cases it is assumed that the sample of elements observed under any treatment combination is a random sample of size n from a potentially infinite population of elements.

Tests of significance in a three-factor experiment follow the same general rules as those for a two-factor experiment. Appropriate F ratios are determined from the structure of the expected values of the mean squares which correspond to the design of the experiment. For the case of model I, all F ratios have MS_{error} for a denominator. For the case of model II given in Table 5.9-2, F ratios for main effects have the following structure:

$$\begin{aligned} H_1: \quad \sigma_a^2 &= 0, & F &= \frac{MS_a}{MS_{\text{error}}}; \\ H_1: \quad \sigma_b^2 &= 0, & F &= \frac{MS_b}{MS_{ab}}; \\ H_1: \quad \sigma_c^2 &= 0, & F &= \frac{MS_c}{MS_{ac}}. \end{aligned}$$

Tests on main effects for model III are considered in Sec. 5.15.

Under model II, the appropriate F ratio to test the hypothesis $\sigma_{\beta\gamma}^2 = 0$ is

$$F = \frac{MS_{bc}}{MS_{abc}}.$$

However, the appropriate F ratio for a test on the hypothesis that $\sigma_{\alpha\beta}^2 = 0$ has the form

$$F = \frac{MS_{ab}}{MS_{error}}.$$

5.10 Simple Effects and Their Tests

The definitions of simple effects will be cast in terms of a three-factor factorial experiment. However, the generalization to higher- or lower-order factorial experiments is direct. Simple effects are associated with both main effects and interaction effects. The former are called simple main effects, the latter simple interaction effects.

A $p \times q \times r$ factorial experiment may be considered from different points of view. For some purposes, it can be considered as a set of qr single-factor experiments on factor A , each experiment being for a different combination of b_j and c_k . For other purposes, it can be considered as a set of r experiments, each experiment being a $p \times q$ factorial experiment for a different level of factor C . It will be convenient, for the present, to adopt this latter point of view. The analysis of the experiment for level c_k takes the following form:

Source of variation	df
A (for c_k)	$p - 1$
B (for c_k)	$q - 1$
AB (for c_k)	$(p - 1)(q - 1)$

The main effects for this single experiment, relative to the entire set of experiments, are called the simple main effects for level c_k . The interaction in this experiment is called the simple AB interaction for level c_k . The variation due to the simple main effects of factor A , $SS_{a \text{ for } c_k}$, is related to the variation of the over-all main effect of factor A and the over-all AC interaction. Specifically,

$$\sum_k SS_{a \text{ for } c_k} = SS_a + SS_{ac}.$$

In words, the sum of the variation due to the simple main effects of A at the various levels of factor C is equal to the variation due to the over-all main effect of factor A and the over-all AC interaction. When the variation due to the AC interaction is zero, the sum of the variation for these simple effects is equal to the over-all main effects. Analogously, it can be shown that

$$\sum_k SS_{ab \text{ for } c_k} = SS_{ab} + SS_{abc}.$$

Table 5.10-1 Definition of Effects

Effect	Definition
Over-all main effect of a_i	$\alpha_i = \mu_{i..} - \mu_{...}$
Simple main effect of a_i for c_k	$\alpha_{i(c_k)} = \mu_{i..k} - \mu_{..k}$
Over-all main effect of b_j	$\beta_j = \mu_{.j.} - \mu_{...}$
Simple main effect of b_j for c_k	$\beta_{j(c_k)} = \mu_{.jk} - \mu_{..k}$
Over-all interaction effect of ab_{ij}	$\alpha\beta_{ji} = \mu_{ij.} - \mu_{...} - \alpha_i - \beta_j$
Simple interaction effect of ab_{ij} for c_k	$\alpha\beta_{ij(c_k)} = \mu_{ijk} - \mu_{..k} - \alpha_{i(c_k)} - \beta_{j(c_k)}$

In words, the sum of the variations of the simple two-factor interactions is equal to the variation of the over-all two-factor interaction plus the three-factor interaction.

The definitions of the simple effects, in terms of population means, are given in Table 5.10-1. It will be noted that simple effects have the same general form as over-all factorial effects; simple effects, however, are restricted to a single level of one or more of the factors. The degree to which over-all effects approximate simple effects depends upon magnitudes of interactions. In the absence of interaction, over-all effects will be equal to corresponding simple effects.

The definition of simple effects will be illustrated by the numerical data given in Table 5.10-2. The data in this table represent population means. An entry in a cell of part i has the general symbol μ_{ijk} ; an entry in a cell of part ii has the general symbol $\mu_{ij.}$; an entry in a cell of part iii has the general

Table 5.10-2 Population Means for $2 \times 2 \times 2$ Factorial Design

(i)

	c_1			c_2		
	b_1	b_2	Mean	b_1	b_2	Mean
	a_1	60	20	40	80	40
a_2	0	0	0	20	20	20
Mean	30	10	20	50	30	40

(ii)

b_1	b_2	Mean	
a_1	70	30	
a_2	10	10	10
Mean	40	20	30

(iii)

c_1	c_2	Mean	
a_1	40	60	
a_2	0	20	10
Mean	20	40	30

(iv)

c_1	c_2	Mean	
b_1	30	50	
b_2	10	30	20
Mean	20	40	30

symbol $\mu_{i.k}$. It is assumed that $P = Q = R = 2$. The simple main effects for factor A at levels c_1 and c_2 are most readily obtained from part iii. It will be found that

$$\begin{aligned}\alpha_1 &= 50 - 30 = 20, & \alpha_2 &= 10 - 30 = -20; \\ \alpha_{1(c_1)} &= 40 - 20 = 20, & \alpha_{2(c_1)} &= 0 - 20 = -20; \\ \alpha_{1(c_2)} &= 60 - 40 = 20, & \alpha_{2(c_2)} &= 20 - 40 = -20.\end{aligned}$$

In each case the simple main effect is equal to the corresponding over-all main effect. This finding indicates that the two-factor interaction effects $\alpha\gamma_{ik}$ are all zero and hence that $\sigma_{\alpha\gamma}^2 = 0$. The $\alpha\gamma_{ik}$'s are most readily obtained from part iii.

$$\begin{aligned}\alpha\gamma_{11} &= 40 - 50 - 20 + 30 = 0, \\ \alpha\gamma_{12} &= 60 - 50 - 40 + 30 = 0, \\ \alpha\gamma_{21} &= 0 - 10 - 20 + 30 = 0, \\ \alpha\gamma_{22} &= 20 - 10 - 40 + 30 = 0.\end{aligned}$$

Conversely, when two-factor interaction effects are found to be zero, corresponding simple main effects will be equal to over-all main effects.

The simple main effects for factor A at levels b_1 and b_2 are most readily obtained from part ii. These over-all and simple main effects are

$$\begin{aligned}\alpha_1 &= 50 - 30 = 20, & \alpha_2 &= 10 - 30 = -20; \\ \alpha_{1(b_1)} &= 70 - 40 = 30, & \alpha_{2(b_1)} &= 10 - 40 = -30; \\ \alpha_{1(b_2)} &= 30 - 20 = 10, & \alpha_{2(b_2)} &= 10 - 20 = -10.\end{aligned}$$

In this case, simple main effects are not equal to corresponding over-all main effects. This finding indicates that the two-factor interactions $\alpha\beta_{ij}$ are not all zero and hence that $\sigma_{\alpha\beta}^2 \neq 0$. The $\alpha\beta_{ij}$'s are obtained from part ii.

$$\begin{aligned}\alpha\beta_{11} &= 70 - 50 - 40 + 30 = 10, \\ \alpha\beta_{12} &= 30 - 50 - 20 + 30 = -10, \\ \alpha\beta_{21} &= 10 - 10 - 40 + 30 = -10, \\ \alpha\beta_{22} &= 10 - 10 - 20 + 30 = 10.\end{aligned}$$

The fact that the $\alpha\beta_{ij}$'s are not all equal to zero implies that the simple main effects are not equal to the corresponding over-all main effects.

The over-all two-factor interaction effects are related to corresponding simple effects. Specifically,

$$(1) \quad \alpha\beta_{ij} = \alpha_{i(b_j)} - \alpha_i.$$

For example,

$$\begin{aligned}\alpha\beta_{12} &= \alpha_{1(b_2)} - \alpha_1 = 10 - 20 = -10, \\ \alpha\beta_{21} &= \alpha_{2(b_1)} - \alpha_2 = -30 - (-20) = -10.\end{aligned}$$

Similarly,

$$(2) \quad \alpha\beta_{ij} = \beta_{j(a_i)} - \beta_j.$$

These last relationships indicate that two-factor interaction effects will be zero when simple main effects are equal to over-all main effects and, conversely, that when two-factor interaction effects are zero corresponding simple main effects will be equal to over-all main effects.

The simple two-factor interaction effects $\alpha\beta_{ij(c_k)}$ are most readily obtained from part i. For example,

$$\begin{aligned}\alpha\beta_{11(c_1)} &= 60 - 40 - 30 + 20 = 10 \\ &= \mu_{111} - \mu_{1.1} - \mu_{.11} + \mu_{..1},\end{aligned}$$

$$\begin{aligned}\alpha\beta_{11(c_2)} &= 80 - 60 - 50 + 40 = 10 \\ &= \mu_{112} - \mu_{1.2} - \mu_{.12} + \mu_{..2}.\end{aligned}$$

Thus it is noted that

$$\alpha\beta_{11(c_k)} = \alpha\beta_{11} = 10, \quad k = 1, 2.$$

In general, the over-all three-factor interaction effect is related to simple two-factor interaction effects by means of the relation

$$(3) \quad \alpha\beta\gamma_{ijk} = \alpha\beta_{ij(c_k)} - \alpha\beta_{ij}.$$

Thus, when the simple interaction effect is equal to the over-all interaction effect, the corresponding three-factor interaction effect will be zero.

For the data in part i in Table 5.10-2,

$$\alpha\beta_{11(c_1)} = \alpha\beta_{11}.$$

Hence from relation (3) it follows that $\alpha\beta\gamma_{111} = 0$. In general, for the data in part i it will be found that

$$\alpha\beta_{ij(c_k)} = \alpha\beta_{ij}, \quad \text{for all } k\text{'s.}$$

From relation (3) it follows that all $\alpha\beta\gamma_{ijk}$'s are equal to zero. Hence $\sigma_{\alpha\beta\gamma}^2 = 0$. Analogous to (3) are the relations

$$(4) \quad \alpha\beta\gamma_{ijk} = \alpha\gamma_{ik(b_j)} - \alpha\gamma_{ik},$$

$$(5) \quad \alpha\beta\gamma_{ijk} = \beta\gamma_{jk(a_i)} - \beta\gamma_{jk}.$$

That relation (3) implies (4) and (5) is illustrated geometrically in the next section.

Estimates of simple effects are in general obtained in the same manner as corresponding over-all effects. For example,

$$\text{est } \alpha_i = \bar{A}_i - \bar{G},$$

whereas

$$\text{est } \alpha_{i(c_k)} = \bar{AC}_{ik} - \bar{C}_k.$$

Similarly,

$$\text{est } \beta_{j(c_k)} = \bar{BC}_{jk} - \bar{C}_k,$$

$$\text{est } \alpha_{i(b_j)} = \bar{AB}_{ij} - \bar{B}_j.$$

The over-all interaction effect is estimated by

$$\text{est } \alpha\beta_{ij} = \bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + G,$$

whereas a simple interaction effect is estimated by

$$\text{est } \alpha\beta_{ij(c_k)} = \overline{ABC}_{ijk} - \overline{AC}_{ik} - \overline{BC}_{jk} + \overline{C}_k.$$

Variation due to simple effects is given by the general formula

$$\Sigma[\text{est (simple effect)}]^2.$$

Computational formulas for such variation are given in Chap. 6.

Tests on simple effects depend upon expected values of mean squares for such effects. In general, appropriate expected values of mean squares for simple effects may be obtained from the expected values appropriate for experiments of lower dimension than the original experiment, i.e., by considering the simple effects as over-all effects of experiments of lower dimension. F ratios are constructed from expected values thus obtained. However, data from the entire experiment may be used to estimate mean squares, where such data are relevant.

5.11 Geometric Interpretation of Higher-order Interactions

Various orders of interactions may be represented geometrically. In terms of such representation, interesting aspects of the meaning of interactions will become clearer. In particular, it will be noted how simple

Table 5.11-1 Population Means

(i)

	c_1					c_2			
	b_1	b_2	b_3			b_1	b_2	b_3	
a_1	20	30	70	40	a_1	0	10	50	20
a_2	60	40	50	50	a_2	40	20	30	30
	40	35	60	45		20	15	40	25

(ii)

	b_1	b_2	b_3	
a_1	10	20	60	30
a_2	50	30	40	40
	30	25	50	35

(iii)

	b_1	b_2	b_3	
c_1	40	35	60	45
c_2	20	15	40	25
	30	25	50	35

interactions are related to over-all interactions. It will also be noted how simple interactions are related to higher-order interactions. The numerical data in Table 5.11-1 will be used for this purpose. It will be assumed that these data represent population means.

Data necessary to plot the profiles for the BC interactions for the two levels of factor A are given in part i. The left-hand side of Fig. 5.1

represents the profiles of means which are in the a_1 row of part i. The dotted line represents the means of the form μ_{1j1} ; the solid line represents means of the form μ_{1j2} . Thus the two profiles on the left represent the BC means for level a_1 . These two profiles have the same shape (i.e., are parallel) This finding implies that all simple interaction effects $\beta\gamma_{jk(a_1)}$ are zero.

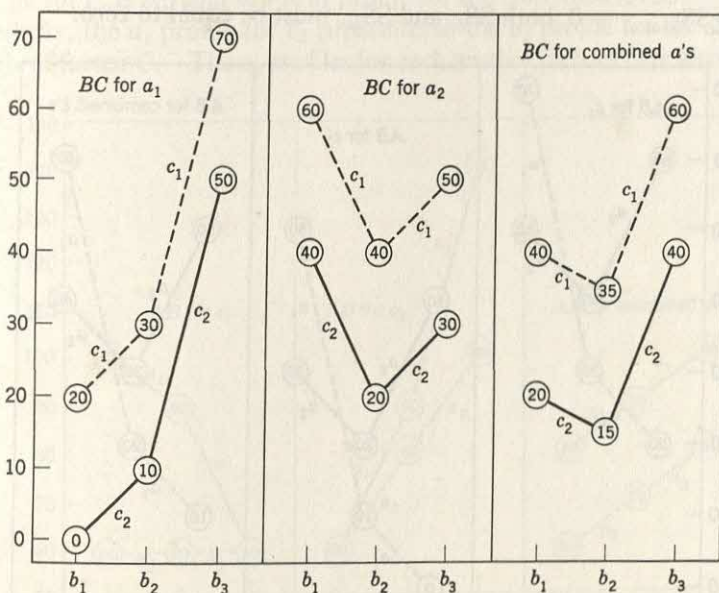


Figure 5.1

In general, the simple two-factor interaction at level a_1 has the definition

$$(1) \quad \beta\gamma_{jk(a_1)} = \mu_{ijk} - \mu_{ij.} - \mu_{i.k} + \mu_{i..}$$

For the data in Table 5.11-1,

$$\beta\gamma_{11(a_1)} = 20 - 10 - 40 + 30 = 0,$$

$$\beta\gamma_{12(a_1)} = 30 - 20 - 40 + 30 = 0,$$

$$\beta\gamma_{13(a_1)} = 70 - 60 - 40 + 30 = 0.$$

The observation that the profiles on the left are parallel has the algebraic counterpart

$$\mu_{1j1} - \mu_{1j2} = \text{constant for all } j\text{'s}.$$

Hence the variation arising from such differences will be zero. This source of variation is that due to the BC interaction at a_1 .

It is noted that the profiles of the BC means for level a_2 are parallel. This finding implies that all $\beta\gamma_{jk(a_2)}$'s are zero. Hence

$$SS_{bc \text{ at } a_1} + SS_{bc \text{ at } a_2} = 0 + 0 = 0.$$

In words, the sum of variations due to the simple interaction effects for the levels of factor A is zero. Since

$$\sum_i SS_{bc \text{ at } a_i} = SS_{bc} + SS_{abc},$$

when $\sum_i SS_{bc \text{ at } a_i} = 0$, both SS_{bc} and SS_{abc} must be equal to zero.

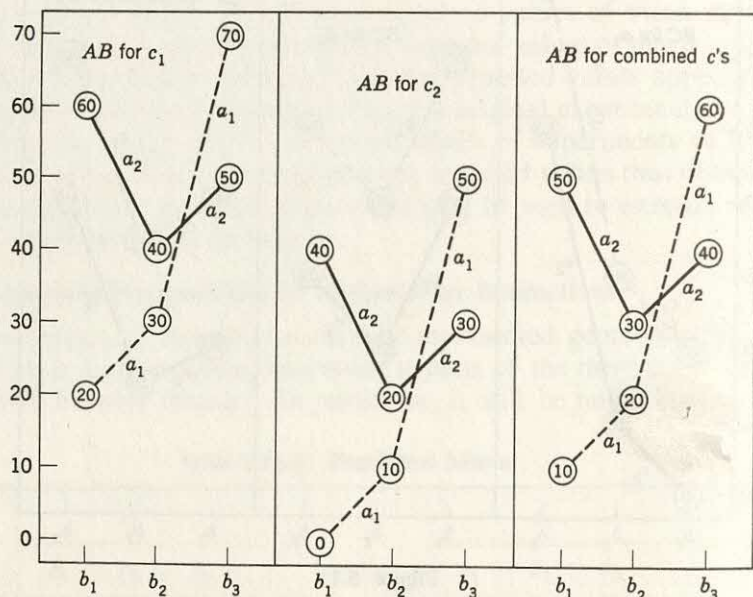


Figure 5.2

The profiles of the over-all BC means, obtained from part iii of Table 5.11-1, are given at the right in Fig. 5.1. The two profiles here are also parallel. This finding implies that

$$\beta\gamma_{jk} = 0, \quad \text{for all } j\text{'s and } k\text{'s.}$$

Hence SS_{bc} must be zero. This latter result was implied by the fact that all the simple interaction effects of the $\beta\gamma_{jk(a_i)}$'s are zero. The fact that the BC profiles for a_1 and a_2 are parallel actually implies that the BC profiles for the combined levels of factor A are also parallel. To summarize, when the simple interactions of two factors at various levels of a third factor are all zero, the corresponding two-factor and three-factor interactions will also be zero.

It is, however, possible for two-factor interactions to be nonzero and yet have zero three-factor interaction. The profiles in Figure 5.2 illustrate this case. These profiles represent the AB means for the two levels of factor

C (Table 5.11-1). The profiles of AB for c_1 are not parallel; the profiles AB at c_2 are not parallel. Hence

$$\sum_k SS_{ab \text{ at } c_k} \neq 0.$$

Although the profiles within each level of factor C are not parallel, the a_1 profile for c_1 is parallel to the a_1 profile for the combined levels of factor C . Similarly, the a_1 profile for c_2 is parallel to the a_1 profile for the combined levels of factor C . The a_2 profiles for each level of factor C are also parallel

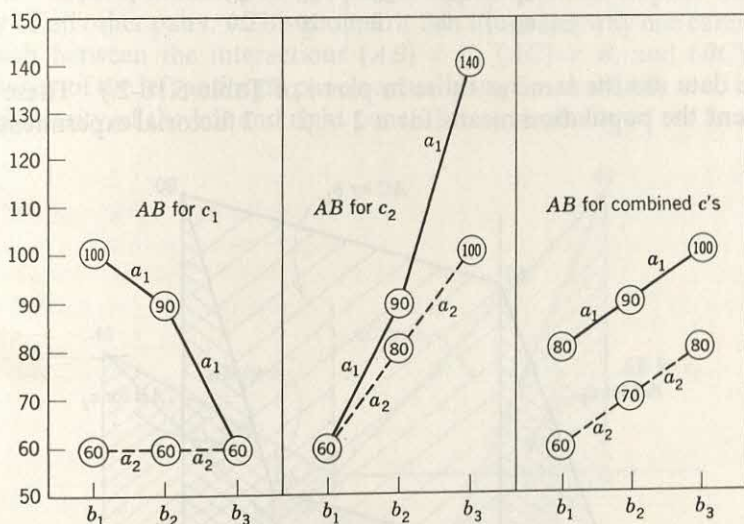


Figure 5.3

to the a_2 profile for the combined data. This finding implies that SS_{abc} will be zero. The fact that the AB profiles for the combined data are not parallel indicates that SS_{ab} will not be zero.

To summarize the implications of the profiles with respect to the three-factor interaction, the latter will be zero when (1) the profiles of the two-factor means are parallel within each level of the third factor or when (2) the pattern of profiles for the two-factor means is geometrically similar to the pattern for the combined levels. In order that patterns be geometrically similar, corresponding profiles must be parallel.

A set of profiles in which the three-factor interaction is nonzero but the two-factor interaction is zero is given in Fig. 5.3. (These profiles are *not* based upon the data in Table 5.11.1.) The profiles within level c_1 are not parallel to each other, nor is the profile at a_1 parallel to the a_1 profile for the combined levels of factor C . Hence the three-factor interaction is nonzero. However, the AB profiles for the combined levels of factor C are parallel. Thus $SS_{ab} = 0$.

The geometric relationships between the simple two-factor interactions, the three-factor interaction, and the over-all two-factor interaction are seen more clearly by drawing three-dimensional profiles. The following data will be used for illustrative purposes.

	c_1		c_2	
	b_1	b_2	b_1	b_2
a_1	60	20	80	40
a_2	0	0	20	20

(These data are the same as those in part i of Table 5.10-2.) These data represent the population means for a $2 \times 2 \times 2$ factorial experiment. A

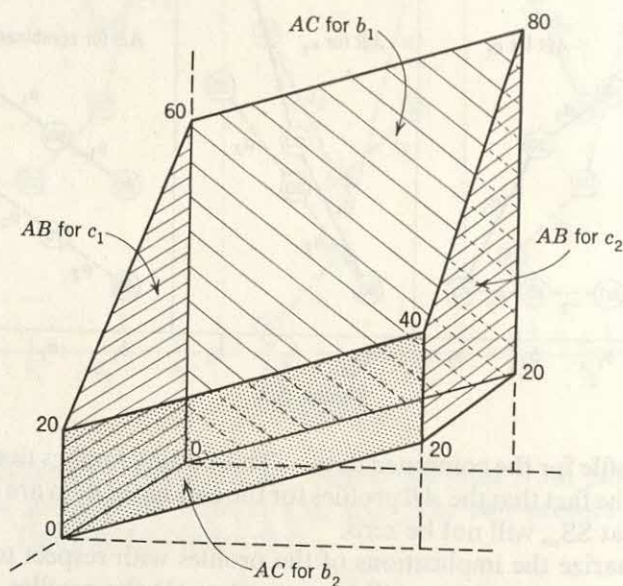


Figure 5.4

geometric representation of the patterns formed by these means is given in Fig. 5.4. The left-hand panel represents the four means in which factor C is at level c_1 . This panel is denoted AB for c_1 . The right-hand panel represents the four means in which factor C is at level c_2 . The line (60,20) in the left panel is parallel to the line (80,40) in the right panel. The line (0,0) in the left panel is parallel to the line (20,20) in the right panel. The vertical lines in these panels are automatically parallel by the method of construction. Since corresponding sides of the left and right panels are parallel, the two panels are geometrically similar.

Geometric similarity of the left and right panels forces similarity of the front and back panels as well as similarity of the top and bottom panels. In analysis-of-variance terms, similarity of panels implies that the profiles of the simple two-factor interactions have the same patterns for all levels of a third factor. Thus the fact that the panel AB for c_1 is similar to the panel AB for c_2 implies that the simple AB interactions have the same pattern for c_1 as they do for c_2 . This in turn implies that $SS_{abc} = 0$.

The fact that $SS_{abc} = 0$ is implied by the similarity of any two opposite panels—in turn, similarity of one pair of opposite panels implies the similarity of all other pairs. This geometric fact illustrates why one cannot distinguish between the interactions $(AB) \times C$, $(AC) \times B$, and $(BC) \times A$. Similarity of the left and right panels actually implies that $(AB) \times C$ is zero. But similarity of the left and right panels forces similarity of the front and

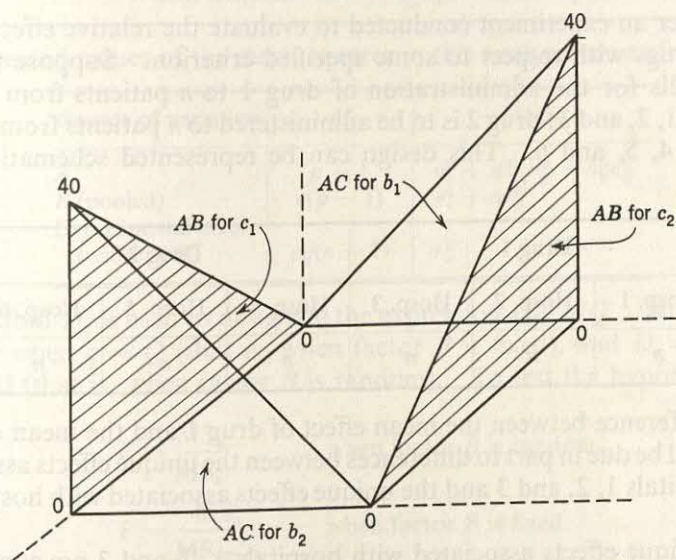


Figure 5.5

back panels. The latter similarity implies that $(AC) \times B$ is zero. Thus, when $(AB) \times C$ is zero, $(AC) \times B$ must also be zero; one implies the other. More emphatically, one is not distinguishable from the other.

When the three-factor interaction is zero, inspection of individual panels will provide information about the two-factor interactions. In Fig. 5.4 the line (60,20) is not parallel to the line (0,0). This implies that the variance due to the AB interaction is not zero. In the front panel, the line (20,40) is parallel to the line (0,20). This implies that the variance due to the AC interaction is zero. In the top panel, the line (60,80) is parallel to the line (20,40). This implies that $SS_{bc} = 0$. When the three-factor interaction is not zero, inspection of the individual panels does not provide

information with respect to the two-factor interactions. The individual panels in pair must be averaged in order to obtain information relevant to the over-all two-factor interaction.

Figure 5.5 illustrates a case in which all main effects are zero, all two-factor interactions are zero, but the three-factor interaction is not zero. Since the left and right panels are not geometrically similar, the simple AB interactions at the two levels of factor C are not zero. Hence the variation due to the three-factor interaction is not zero. When corresponding points on the left and right panels are averaged, the top line of the average profile is (20,20) and the bottom line is (0,0). Hence in the over-all AB profile the top and bottom line are parallel; thus the variation due to the over-all AB interaction is zero.

5.12 Nested Factors (Hierarchal Designs)

Consider an experiment conducted to evaluate the relative effectiveness of two drugs with respect to some specified criterion. Suppose that the design calls for the administration of drug 1 to n patients from each of hospitals 1, 2, and 3; drug 2 is to be administered to n patients from each of hospitals 4, 5, and 6. This design can be represented schematically as follows:

Drug 1			Drug 2		
Hosp. 1	Hosp. 2	Hosp. 3	Hosp. 4	Hosp. 5	Hosp. 6
n	n	n	n	n	n

The difference between the mean effect of drug 1 and the mean effect of drug 2 will be due in part to differences between the unique effects associated with hospitals 1, 2, and 3 and the unique effects associated with hospitals 4, 5, and 6.

The unique effects associated with hospitals 1, 2, and 3 are confined to drug 1 whereas the unique effects associated with hospitals 4, 5, and 6 are confined to drug 2. Effects which are restricted to a single level of a factor are said to be *nested* within that factor. In the experimental design being considered, the hospital effects are nested under the drug factor. Since a given hospital appears only under one of the two drugs, there is no way of evaluating the interaction effect between the hospital and the drug. Before such an interaction effect can be evaluated, each hospital must appear under both levels of the drug factor.

Thus, in a two-factor experiment having one factor nested under the other, the interaction effect cannot be evaluated. For the general case of a two-factor experiment in which factor B is nested under factor A , the structural model is

$$\overline{AB}_{ij} = \mu_{..} + \alpha_i + \beta_{j(i)} + \bar{\epsilon}_{ij}.$$

The notation $\beta_{j(i)}$ indicates that the effect of level b_j is nested under level a_i . Note that no interaction term of the form $\alpha\beta_{ij(i)}$ appears in the model. Inferences made from this type of design assume implicitly that the variation associated with this latter interaction is either zero or negligible relative to the variation associated with the main effects.

The analysis of variance for the design outlined at the beginning of this section takes the following form:

Factor		df	df for general case
<i>A</i>	Drug	1	$p - 1$
<i>B</i> (w. a_1)	Hospitals within drug 1	2	$q - 1$
<i>B</i> (w. a_2)	Hospitals within drug 2	2	$q - 1$
	Within hospital	$6(n - 1)$	$pq(n - 1)$

The expected values of the mean squares in this analysis are as follows:

Source of variation	df	E(MS)
<i>A</i>	$p - 1$	$\sigma_e^2 + nD_q\sigma_\beta^2 + nq\sigma_\alpha^2$
<i>B</i> (pooled)	$q(p - 1)$	$\sigma_e^2 + n\sigma_\beta^2$
Experimental error (within cell)	$pq(n - 1)$	σ_e^2

The symbol D_q is used to designate the expression $1 - q/Q$. Numerically $D_q = 0$ when $q = Q$ (that is, when factor B is fixed), and $D_q = 1$ when $q/Q = 0$ (that is, when factor B is random). To test the hypothesis that $\sigma_\alpha^2 = 0$,

$$F = \frac{MS_a}{MS_b}, \quad \text{when factor } B \text{ is random;}$$

$$F = \frac{MS_a}{MS_{\text{error}}}, \quad \text{when factor } B \text{ is fixed.}$$

By way of contrast, in a two-factor factorial experiment each level of one of the factors is associated with each level of the second factor. If the design outlined at the beginning of this section were changed to a two-factor factorial experiment, the new design could be represented schematically as follows:

	Hosp. 1	Hosp. 2	Hosp. 3	Hosp. 4	Hosp. 5	Hosp. 6
Drug 1	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$
Drug 2	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$	$n/2$

This factorial experiment requires n subjects from each of the hospitals, but $n/2$ of the subjects from each hospital are given drug 1, and $n/2$ of the

subjects are given drug 2. In many cases the two-factor factorial experiment is to be preferred to the two-factor design in which the hospital factor is nested under the drug factor—particularly in those cases in which an interaction effect might be suspected. However, there are some instances in which the experimenter may be forced to use a design in which one factor is nested within another.

As another illustration of nested effects, consider the following experimental design:

	Drug 1			Drug 2		
	Hosp. 1	Hosp. 2	Hosp. 3	Hosp. 4	Hosp. 5	Hosp. 6
Category 1	n	n	n	n	n	n
Category 2	n	n	n	n	n	n

This design calls for a sample of n patients in category 1 and n patients in category 2 (random samples) from each of the hospitals. Patients from hospitals 1, 2, and 3 receive drug 1; patients from hospitals 4, 5, and 6 receive drug 2. In this design the hospital factor is nested under the drug

Table 5.12-1 Analysis of Three-factor Experiment in Which Factor B is Nested under Factor A

Source of variation		df	df for general case
A	Drugs	1	$p - 1$
$B(w. a_1)$	Hospital w. drug 1	2	$q - 1$
.	.	.	.
.	.	.	.
.	.	4	$p(q - 1)$
$B(w. a_p)$	Hospital w. drug 2	2	$q - 1$
C	Categories	1	$r - 1$
AC	Drug \times category	1	$(p - 1)(r - 1)$
$(B w. A) \times C$	Hospital \times category	4	$p(q - 1)(r - 1)$
Error	Within cell	$12(n - 1)$	$pqr(n - 1)$
Total		$11 + 12(n - 1)$	$npqr - 1$

factor. Since some patients from each category receive drug 1 and some patients from each category receive drug 2, the category factor is not nested under the drug factor. Further, since patients from each of the categories are obtained from each hospital, the category factor is not nested under the hospital factor.

The design that has just been sketched may be considered as a three-factor experiment, the factors being drugs (A), hospitals (B), and categories (C). In this case, factor B is nested under factor A , but all other relationships are

those of a factorial experiment. The model for this type of experiment has the following form:

$$\overline{ABC}_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \bar{\epsilon}_{ijk}.$$

No interaction in which the subscript i appears twice occurs in this model. That is, the interactions $\alpha\beta_{ij(i)}$ and $\alpha\beta\gamma_{ij(i)k}$ do not appear. Hence the utility of this type of design is limited to situations in which such interactions are either zero or negligible relative to other sources of variation of interest.

The analysis of variance for this design takes the form given in Table 5.12-1. In making tests by means of the F ratio, the appropriate ratios are determined from the expected values of the mean squares. The latter are

Table 5.12-2 Expected Values of Mean Squares, Factor B Nested under Factor A

Source of variation	df	Expected value of mean square
A	$p - 1$	$\sigma_e^2 + nD_qD_r\sigma_{\beta\gamma}^2 + nqD_r\sigma_{\alpha\gamma}^2 + nrD_q\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
$B \text{ w. } A$	$p(q - 1)$	$\sigma_e^2 + nD_r\sigma_{\beta\gamma}^2 + nr\sigma_{\beta}^2$
C	$r - 1$	$\sigma_e^2 + nD_q\sigma_{\beta\gamma}^2 + nqD_p\sigma_{\alpha\gamma}^2 + npq\sigma_{\gamma}^2$
AC	$(p - 1)(r - 1)$	$\sigma_e^2 + nD_q\sigma_{\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
$(B \text{ w. } A)C$	$p(q - 1)(r - 1)$	$\sigma_e^2 + n\sigma_{\beta\gamma}^2$
Within cell	$pqr(n - 1)$	σ_e^2

in part a function of whether the factors in the experiment are fixed or random. General expressions for the expected values of mean square appropriate for this type of design are given in Table 5.12-2. In this table the symbol D_p is used to designate $1 - p/P$, D_q is used to designate $1 - q/Q$, and D_r is used to designate $1 - r/R$. Each of these D 's is either 0 or 1 depending, respectively, on whether the corresponding factor is fixed or random.

In this last table, the source of variation $B \text{ w. } A$ (the main effects due to factor B , which is nested within factor A) is actually the sum of the following main effects:

$$\begin{array}{rcl} & & \text{df} \\ B \text{ w. } a_1 & & q - 1 \\ \dots & & \dots \\ B \text{ w. } a_p & & q - 1 \\ \text{Sum} = B \text{ w. } A & & p(q - 1) \end{array}$$

Similarly, the interaction $(B \text{ w. } A) \times C$ represents a pooling of the following interactions:

$$\begin{array}{rcl} & & \text{df} \\ (B \text{ w. } a_1)C & & (q - 1)(r - 1) \\ \dots & & \dots \\ (B \text{ w. } a_p)C & & (q - 1)(r - 1) \\ \text{Sum} = (B \text{ w. } A)C & & p(q - 1)(r - 1) \end{array}$$

If this pooled interaction is used in the denominator of an F ratio, the variations which are pooled must be homogeneous if the resulting F ratio is to have an F distribution when the hypothesis being tested is true.

As another illustration of nested effects, consider an experiment in which the subjects have been classified as follows:

Company 1				Company 2			
Dept. 1		Dept. 2		Dept. 3		Dept. 4	
Job 1	Job 2	Job 3	Job 4	Job 5	Job 6	Job 7	Job 8
n	n	n	n	n	n	n	n

Suppose that n people from each job are included in an experiment in which attitude toward a retirement plan is being studied. This design may be considered as a three-factor experiment in which the department factor (B) is nested under the company factor (A). The job factor (C) is nested under both factors B and A . This type of design is referred to as a *hierarchal* design. In a three-factor hierarchal experiment, factor B is nested under factor A , and factor C is nested under both factors B and A . (The design in Table 5.12-1 is only partially hierarchal, since factor C was not nested under either factor A or factor B .)

The model for a three-factor hierarchal experiment has the form

$$\overline{ABC}_{ijk} = \mu_{...} + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \varepsilon_{ijk}.$$

The notation $\gamma_{k(ij)}$ indicates that factor C is nested under both factors A and B . It should be noted that no interaction terms appear explicitly in this model. The expected values of the mean squares for this design are summarized in Table 5.12-3. The numerical values of the D 's in these expected values depend upon the respective sampling fractions for the levels of the factors.

Table 5.12-3 Expected Values of Mean Squares for Three-factor Hierarchal Experiment

Source of variation	df	Expected value of mean square
A	$p - 1$	$\sigma_e^2 + nD_r\sigma_\gamma^2 + nrD_q\sigma_\beta^2 + nqr\sigma_\alpha^2$
B w. A	$p(q - 1)$	$\sigma_e^2 + nD_r\sigma_\gamma^2 + nr\sigma_\beta^2$
C w. (A and B)	$pq(r - 1)$	$\sigma_e^2 + n\sigma_\gamma^2$
Experimental error	$pqr(n - 1)$	σ_e^2
Total	$npqr - 1$	

The expected values of the mean squares for a completely hierarchal design are readily obtained, once the expected value of the mean square for the factor within which all other factors are nested is determined. The

expected value for each succeeding factor is obtained from the one above by dropping the last term and making D in the next to the last term unity.

The following design represents a partially hierarchical design:

	City 1		City 2	
	School 1	School 2	School 3	School 4
Method 1	n	n	n	n
Method 2	n	n	n	n

Suppose that the purpose of this experiment is to evaluate the relative effectiveness of two different methods of teaching a specified course. Since both methods of training are given within each of the schools and within each of the cities, there is no nesting with respect to the methods factor. The school factor is, however, nested within the city factor. The expected values of the mean squares for this design have the form given in Table 5.12-4. This

Table 5.12-4 Expected Values of Mean Squares for Three-factor Partially Hierarchical Experiment

Source of variation		df	Expected value of mean square
A	Methods	$p - 1$	$\sigma_e^2 + nD_r\sigma_{\alpha\gamma}^2 + nrD_q\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
B	Cities	$q - 1$	$\sigma_e^2 + nD_rD_p\sigma_{\alpha\gamma}^2 + nrD_p\sigma_{\alpha\beta}^2 + npD_r\sigma_{\gamma}^2 + npr\sigma_{\beta}^2$
$C \text{ w. } B$	Schools within cities	$q(r - 1)$	$\sigma_e^2 + nD_p\sigma_{\alpha\gamma}^2 + np\sigma_{\gamma}^2$
AB		$(p - 1)(q - 1)$	$\sigma_e^2 + nD_r\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2$
$A \times (C \text{ w. } B)$		$q(p - 1)(r - 1)$	$\sigma_e^2 + n\sigma_{\alpha\gamma}^2$
	Within cell	$pqr(n - 1)$	σ_e^2

design enables the experimenter to eliminate systematic sources of variation associated with differences between cities and differences between schools within cities from the experimental error—at the cost, however, of reduced degrees of freedom for experimental error.

As still another example of a useful design involving a nested factor, consider an experiment with the following schematic representation:

	Method 1			Method 2		
	Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
Period 1						
Period 2						
Period 3						
Period 4						

In this design persons 1, 2, and 3 are observed under method of training 1; criterion measures are obtained at four different periods during the training process. Persons 4, 5, and 6 are observed at comparable periods under training method 2. This design may be considered as a three-factor experiment in which the person factor is nested under the methods factor. If methods is considered to be factor A , persons factor B , and periods factor C , the model for this design is

$$ABC_{ijk} = \mu_{..} + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{j(i)k} + \varepsilon_{ijk}.$$

For this design there is only one observation in each cell of the experiment. Hence there is no within-cell variation. The analysis of variance is identical in form with that given in Table 5.12-2 if the within-cell variation is deleted. The expected values of the mean squares are also identical to those given in

Table 5.12-5 Three-factor Partially Hierarchal Design
(Factors A and C fixed, B random)

Source of variation	df	Expected value of mean square
A Methods	$p - 1$	$\sigma_{\varepsilon}^2 + r\sigma_{\beta}^2 + qr\sigma_{\alpha}^2$
B w. A People within methods	$p(q - 1)$	$\sigma_{\varepsilon}^2 + r\sigma_{\beta}^2$
C Periods	$r - 1$	$\sigma_{\varepsilon}^2 + \sigma_{\beta\gamma}^2 + pq\sigma_{\gamma}^2$
AC Method \times period	$(p - 1)(r - 1)$	$\sigma_{\varepsilon}^2 + \sigma_{\beta\gamma}^2 + q\sigma_{\alpha\gamma}^2$
$(B$ w. $A)C$	$p(q - 1)(r - 1)$	$\sigma_{\varepsilon}^2 + \sigma_{\beta\gamma}^2$

Table 5.12-2, with n set equal to unity. Specializing this design to the case in which factors A (methods) and C (periods) are fixed and factor B (persons) is random, one obtains the expected values given in Table 5.12-5. In this table, the source of variation due to B w. A represents the pooled variation of people within methods. A homogeneity assumption is required for this pooling. The interaction term $(B$ w. $A)C$ also represents a pooling of different sources of variation. The homogeneity assumption required for pooling in this case is equivalent to the assumption that the correlation between periods be constant within each of the methods. The F tests for the analysis in Table 5.12-5 have the following form:

$$H_1: \sigma_{\alpha}^2 = 0, \quad F = \frac{MS_{\text{methods}}}{MS_{\text{people w. methods}}};$$

$$H_1: \sigma_{\gamma}^2 = 0, \quad F = \frac{MS_{\text{periods}}}{MS_{(B \text{ w. } A)C}};$$

$$H_1: \sigma_{\alpha\gamma}^2 = 0, \quad F = \frac{MS_{\text{method} \times \text{period}}}{MS_{(B \text{ w. } A)C}}.$$

Rather than considering this last experiment as a three-factor partially hierarchal experiment, there are other means for classifying this type of experiment. In particular, it can be considered as a two-factor one in

which there are repeated measurements on one of the factors. This latter type of classification will receive more extensive treatment in Chap. 7.

5.13 Split-plot Designs

The *split-plot* design has much in common with partially hierarchical designs. The term split-plot comes from agricultural experimentation in which a single level of one treatment is applied to a relatively large plot of ground (the whole plot) but all levels of a second treatment are applied to subplots within the whole plot. For example, consider the following design, in which the levels of factor A are applied to the whole plots and the levels of factor B are applied to the subplots:

a_1 Plot 1	a_2 Plot 2	a_1 Plot 3	a_2 Plot 4
b_2	b_1	b_2	b_3
b_3	b_3	b_1	b_2
b_1	b_2	b_3	b_1

In this design, differences between the levels of factor A cannot be estimated independently of differences between groups of blocks. That is,

$$\begin{aligned} \text{est}(\alpha_1 - \alpha_2) &= (\text{plot 1} + \text{plot 3}) - (\text{plot 2} + \text{plot 4}) \\ &= A_1 - A_2. \end{aligned}$$

For this reason the variation due to the levels of factor A is part of the between-plot effects. However, comparisons among the levels of factor B are part of the within-plot variation. From the information within each of the plots, estimates of the main effects due to factor B may be obtained. These estimates are free of variation due to whole plots. In the analysis of this type of design, sources which are part of the whole-plot variation are usually grouped separately from those which are part of the within-plot variation.

The model for this experiment is

$$X_{ijk} = \mu_{...} + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \pi'_{j(ik)} + \varepsilon_{ijk}$$

The notation $\pi_{k(i)}$ designates the effect of plot k within level a_i . (This notation indicates that the plot effects are nested within the levels of factor A .) The notation $\pi'_{j(ik)}$ designates residual subplot effects. For the special, but frequently occurring, case in which A and B are fixed factors and the plots are a random sample from a specified population of plots, the analysis of variance assumes the form given in Table 5.13-1. In this analysis, each of the p levels of factor A is assigned at random to np plots. Within each plot, the levels of factor B are assigned at random to the subplots. The expected values of the mean squares in this table are actually identical in

form with those given in Table 5.12-5 if plots are assigned the role of factor B in this latter design.

There is a distinction between the designs usually placed in the hierarchal category and designs in the split-plot category. In the hierarchal designs, generally (but not always) all except one of the factors are modes of classifying the experimental units rather than treatments which are administered to the units by the experimenter. Such modes of classification are set up primarily to eliminate, in part, differences among the experimental units

Table 5.13-1 Expected Values of Mean Squares for Split-plot Design (A and B fixed, plots random)

Source of variation	df	Expected value of mean square
Between plots	$np - 1$	
A	$p - 1$	$\sigma^2 + q\sigma_\pi^2 + nq\sigma_\alpha^2$
Plots w. a_1	$p(n - 1)$	$\sigma_\epsilon^2 + q\sigma_\pi^2$
Plots w. a_p		
Within plots	$np(q - 1)$	
B	$q - 1$	$\sigma_\epsilon^2 + \sigma_\pi^2 + np\sigma_\beta^2$
AB	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + \sigma_\pi^2 + n\sigma_{\alpha\beta}^2$
$B \times$ plots w. a_1	$p(q - 1)(n - 1)$	$\sigma_\epsilon^2 + \sigma_\pi^2$
Plots w. a_p		

from the experimental error. Interaction between the treatment and the classifications is generally considered to be negligible. The usual hierarchal experiment may be regarded as a single-factor experiment with controls on the grouping of the experimental units.

In contrast, the split-plot design has two treatment factors. Whole plots are the experimental units for one of the factors, whereas the subplots are the experimental units for the second factor. If this distinction is made, the final design considered in the last section should be placed in the split-plot rather than the partially hierarchal category.

By inspection of the design outlined at the beginning of this section it is not obvious that the AB interaction is free of variation due to whole-plot effects. To demonstrate that such is the case, consider only the whole-plot effects in each of the following means:

Mean	Whole-plot effects
\bar{A}_1	$\pi_1 + \pi_3$
\bar{B}_1	$\pi_1 + \pi_2 + \pi_3 + \pi_4$
\overline{AB}_{11}	$\pi_1 + \pi_3$
\bar{G}	$\pi_1 + \pi_2 + \pi_3 + \pi_4$

An estimate of the interaction effect associated with treatment combination ab_{11} is

$$\text{est } \alpha\beta_{11} = \overline{AB}_{11} - \bar{A}_1 - \bar{B}_1 + \bar{G}.$$

The whole-plot effects associated with $\text{est } \alpha\beta_{11}$ are

$$(\pi_1 + \pi_3) - (\pi_1 + \pi_3) - (\pi_1 + \pi_2 + \pi_3 + \pi_4) + (\pi_1 + \pi_2 + \pi_3 + \pi_4) = 0.$$

Thus the whole-plot effects for this interaction sum to zero. In general, the whole-plot effects for $\text{est } \alpha\beta_{ij}$ will sum to zero for all i 's and j 's. Hence the variation due to whole-plot effects does not influence variation due to the AB interaction.

It should be noted, however, that the variation due to the AB interaction is not free of variation, if any, associated with the $B \times \text{plot}$ interaction. This latter source of variation may be analyzed as follows:

$B \times \text{plot}$	$q(pr - 1)$
AB	$(p - 1)(r - 1)$
$B \times \text{plots w. } a_1$	$p(q - 1)(r - 1)$
...	
$B \times \text{plots w. } a_p$	

This analysis of the $B \times \text{plot}$ interaction shows the AB interaction as part of this source of variation. This actually follows directly from the fact that the effects of factor A are confounded with groups of plots.

Several variations of the split-plot design are possible. One such variation is a double-split or split-split-plot design. In this type of design each level of factor A is assigned at random to n whole plots. (A total of np whole plots is required for this design.) Each of the whole plots is divided into q subplots. The q levels of factor B are then assigned at random to the subplots within each whole plot, and each subplot is divided into r sub-subplots. The r levels of factor C are then assigned at random to each of the sub-subplots. Part of a split-split-plot design is illustrated schematically below:

a_1	a_2
bc_{11}	bc_{22}
.....
bc_{13}	bc_{21}
.....
bc_{12}	bc_{23}
bc_{22}	bc_{13}
.....
bc_{21}	bc_{12}
.....
bc_{23}	bc_{11}

Thus the experimental unit for factor A is the whole plot; the experimental unit for factor B is the subplot; and the experimental unit for factor C is the sub-subplot. Since the sub-subplots are nested within the subplots and the latter are nested within the whole plots, factor C is nested under the subplots and factor B is nested under the whole plots. Factor A is partially confounded with groups of whole plots.

The model for this type of design may be written

$$X_{ijkm} = \mu_{...} + \alpha_i + \pi_{m(i)} + \beta_j + \alpha\beta_{ij} + \pi'_{m(ij)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} \\ + \pi''_{m(ijk)} + \varepsilon_{ijkm}.$$

The notation $\pi''_{m(ij)}$ designates the residual sub-subplot effect. (This latter may also be regarded as the pooled $\gamma\pi_{km(i)}$ and $\beta\gamma\pi_{jkm(i)}$ interaction effects.) The analysis of this type of experiment takes the form given in Table 5.13-2.

Table 5.13-2 Expected Values of Mean Squares for Split-split-plot Design
(A, B, C fixed, plots random)

Source of variation	df	Expected value of mean square
Between whole plots	$np - 1$	
A	$p - 1$	$\sigma_\varepsilon^2 + qr\sigma_\pi^2 + nqr\sigma_\alpha^2$
Whole-plot residual	$p(n - 1)$	$\sigma_\varepsilon^2 + qr\sigma_\pi^2$
Within subplots	$np(q - 1)$	
B	$q - 1$	$\sigma_\varepsilon^2 + r\sigma_{\pi'}^2 + npr\sigma_\beta^2$
AB	$(p - 1)(q - 1)$	$\sigma_\varepsilon^2 + r\sigma_{\pi'}^2 + nr\sigma_{\alpha\beta}^2$
Subplot residual	$p(n - 1)(q - 1)$	$\sigma_\varepsilon^2 + r\sigma_{\pi'}^2$
Within sub-subplots	$npq(r - 1)$	
C	$r - 1$	$\sigma_\varepsilon^2 + \sigma_{\pi''}^2 + npq\sigma_\gamma^2$
AC	$(p - 1)(r - 1)$	$\sigma_\varepsilon^2 + \sigma_{\pi''}^2 + nq\sigma_{\alpha\gamma}^2$
BC	$(q - 1)(r - 1)$	$\sigma_\varepsilon^2 + \sigma_{\pi''}^2 + np\sigma_{\beta\gamma}^2$
ABC	$(p - 1)(q - 1)(r - 1)$	$\sigma_\varepsilon^2 + \sigma_{\pi''}^2 + n\sigma_{\alpha\beta\gamma}^2$
Sub-subplot residual	$pq(n - 1)(r - 1)$	$\sigma_\varepsilon^2 + \sigma_{\pi''}^2$

The expected values of the mean squares given in this table are for the special case in which factors A, B , and C are considered fixed and plots are considered random. The numbers of levels of factor A, B , and C are, respectively, p, q , and r ; each level of factor A is assigned to n whole plots.

The expected values of the mean squares indicate the structure of appropriate F ratios. It will be noted that the sub-subplot residual mean square is the appropriate denominator for the main effect and all interactions involving factor C . More extensive consideration of split-plot designs is given in connection with designs having repeated measures on the same people. These designs form the subject matter of Chap. 7.

The design summarized in Table 5.13-2 involves pqr cells. Within each cell there are n independent observations. Hence the within-cell variation has $pqr(n - 1)$ degrees of freedom. It is of interest to note that the sum of the degrees of freedom for the residuals is the degrees of freedom for the within-cell variation.

Source	df
Whole-plot residual	$p(n - 1)$
Subplot residual	$p(n - 1)(q - 1)$
Sub-subplot residual	$pq(n - 1)(r - 1)$
Total	$pqr(n - 1)$

5.14 Rules for Deriving the Expected Values of Mean Squares

Given an experiment in which the underlying variables can be assumed to satisfy the conditions of the general linear model, the expected values of the mean squares computed from the experimental data can be obtained by means of a relatively simple set of rules. Although these rules lead to an end product which has been proved to be statistically correct when the assumptions underlying the general linear model are met, the rules themselves provide little insight into the mathematical rationale underlying the end product. The assumptions that underlie the general linear model under consideration have been stated in Sec. 5.6. The rules which will be outlined in this section are those developed by Cornfield and Tukey (1956). A similar set of rules will also be found in Bennett and Franklin (1954).

As the rules will be given, no distinction will be made between random and fixed factors. However, certain of the terms become either 0 or 1 depending upon whether an experimental variable corresponds to a fixed or random factor. For purposes of simplifying the expressions that will result, the following notation will be used:

$$\begin{aligned} D_p &= 1 - p/P, & i &= 1, \dots, p. \\ D_q &= 1 - q/Q, & j &= 1, \dots, q. \\ D_r &= 1 - r/R, & k &= 1, \dots, r. \end{aligned}$$

If $p = P$, that is, if factor A is fixed, then D_p is zero. On the other hand, if factor A is random, D_p is unity. Similarly, the other D 's are either 0 or 1 depending upon whether the corresponding factor is fixed or random. In the application of these rules to designs of special interest, the appropriate evaluation of the D 's should be used rather than the D 's themselves. The general statement of the rules is followed by a series of examples.

Rule 1. Write the appropriate model for the design, making explicit in the notation those effects which are nested.

Rule 2. Construct a two-way table in which the terms in the model (except the grand mean) are the row headings and the subscripts appearing in the model are the column headings. The number of columns in this table will be equal to the number of different subscripts in the model. The number

of rows will be equal to the number of terms in the model which have subscripts. The row headings should include all subscripts associated with a given term in the model.

Rule 3. To obtain the entries in column i ,

- enter D_i in those rows having headings containing an i which is not nested,
- enter unity in those rows having headings containing an i which is nested,
- enter p in those rows having headings which do not contain an i .

Rule 4. To obtain entries in column j ,

- enter D_j in those rows having headings containing a j which is not nested,
- enter unity in those rows having headings containing a j which is nested,
- enter q in those rows having headings which do not contain a j .

Rule 5. Entries in all other columns follow the general pattern outlined in rules 3 and 4. For example, the possible entries in column k would be D_r , unity, and r .

Rule 6. The expected value of the mean square for the main effect of factor A is a weighted sum of the variances due to all effects which contain the subscript i . If a row heading contains a subscript i , then the weight for the variance due to this row effect is the product of all entries in this row, the entry in column i being omitted. (For nested effects, see rule 10.)

Rule 7. The expected value of the mean square for the main effect of factor B is a weighted sum of the variances due to all effects which contain the subscript j . If a row heading contains a j , the weight for the variance due to this effect is the product of all entries in this row, the entry in column j being omitted. (See rule 10.)

Rule 8. The expected value of the mean square for the AB interaction is a weighted sum of the variances due to all effects which contain both the subscripts i and j . If a row heading contains both the subscripts i and j , then the weight for the variance corresponding to this effect is the product of all entries in this row, the entries in *both* columns i and j being omitted.

Rule 9. In general, the expected value of a mean square for an effect which has the general representation XYZ_{uvw} is a weighted sum of the variances due to all effects in the model which contain all the subscripts u , v , and w (and possibly other subscripts). If a row heading does contain all three of the three subscripts u , v , and w , then the weight for the variance due to the corresponding row effects is the product of all entries in this row, the entries in columns u , v , and w being omitted.

Rule 10. If an effect is nested, the expected value of its mean square is a weighted sum of variances corresponding to all effects containing the same

subscripts as the nested effect. For example, if the main effect of factor B appears as $\beta_{j(i)}$ in the model, then the relevant effects are those which contain both the subscripts i and j . Similarly, if the term $\beta\gamma_{j(i)k}$ appears in the model, in considering the set of relevant variances, row headings must contain all three of the subscripts i, j , and k .

The application of these rules will be illustrated by means of a $p \times q \times r$ partially hierarchal factorial design having n observations per cell. In this design, it will be assumed that factor B is nested under factor A . This type of design may be represented schematically as follows:

	a_1			a_2		
	$b_{1(1)}$	$b_{2(1)}$	$b_{3(1)}$	$b_{1(2)}$	$b_{2(2)}$	$b_{3(2)}$
c_1	n observations in each cell					
c_2						
c_3						
c_4						

The notation $b_{j(i)}$ indicates that factor B is nested under factor A . The structural model for this design may be written as

$$X_{ijkm} = \mu_{...} + \alpha_i + \beta_{j(i)} + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{j(i)k} + \varepsilon_{m(ijk)}.$$

In accordance with rule 1, the notation in the structural model makes explicit those effects which are nested. Thus, the notation $\beta_{j(i)}$ indicates that the levels of factor B are nested under the levels of factor A . The notation $\varepsilon_{m(ijk)}$ indicates that the unique effects associated with an observation on element m in cell abc_{ijk} are nested under all effects; i.e., the experimental error is nested under all factors.

Table 5.14-1 Expected Value of Mean Squares for Partially Hierarchal Factorial Design

Effect	i	j	k	m	E(MS)
α_i	D_p	q	r	n	$\sigma_e^2 + nD_qD_r\sigma_{\beta\gamma}^2 + nqD_r\sigma_{\alpha\gamma}^2 + nrD_q\sigma_{\beta}^2 + nqr\sigma_{\alpha}^2$
$\beta_{j(i)}$	1	D_q	r	n	$\sigma_e^2 + nD_r\sigma_{\beta\gamma}^2 + nr\sigma_{\beta}^2$
γ_k	p	q	D_r	n	$\sigma_e^2 + nD_q\sigma_{\beta\gamma}^2 + nqD_p\sigma_{\alpha\gamma}^2 + np\sigma_{\gamma}^2$
$\alpha\gamma_{ik}$	D_p	q	D_r	n	$\sigma_e^2 + nD_q\sigma_{\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
$\beta\gamma_{j(i)k}$	1	D_q	D_r	n	$\sigma_e^2 + n\sigma_{\beta\gamma}^2$
$\varepsilon_{m(ijk)}$	1	1	1	1	σ_e^2

The two-way table called for by rule 2 is given in Table 5.14-1. The row headings are the terms in the model, the term $\mu_{...}$ being omitted. The column headings are the different subscripts that appear in the model.

The entries that appear in column i of this table were obtained in accordance with rule 3. Since rows α_i and $\alpha\gamma_{ik}$ contain a subscript i which is not in parentheses, the entry in column i for each of these rows is D_p . Since rows

$\beta_{j(i)}$, $\beta\gamma_{j(i)k}$, and $\varepsilon_{m(ijk)}$ each contains the subscript i in parentheses, the entry in column i for each of these rows is unity. All other entries in column i are p , since none of the remaining row headings contains the subscript i .

The entries that appear in column j were obtained in accordance with rule 4. The entries that appear in column m follow from rule 5—with one exception. From rule 5 the entry in row $\varepsilon_{m(ijk)}$ would be D_n . Since the experimental error will always be considered a random variable, $D_n = 1 - (n/N)$ will always be considered equal to unity. Thus the n observations that appear within a cell in this design are considered to be a random sample from a potentially infinite number of observations that could be made within a cell.

In accordance with rule 6, the expected value of the mean square for the main effect of factor A is a weighted sum of the variances due to all row effects which contain the subscript i . Thus $E(MS_a)$ is a weighted sum of the following variances:

$$\sigma_e^2, \sigma_{\beta\gamma}^2, \sigma_{\alpha\gamma}^2, \sigma_\beta^2, \sigma_\alpha^2.$$

The weight for σ_α^2 , which is given by the product of all terms in row α_i , the term in column i being omitted, is nqr . The weight for σ_β^2 , which is the product of all terms in row $\beta_{j(i)}$, the term in column i being omitted, turns out to be nrD_a . The weight for $\sigma_{\alpha\gamma}^2$ is the product of all terms in row $\alpha\gamma_{ik}$, the term in column i being omitted; this weight is nqD_r . The properly weighted sum for $E(MS_a)$ appears to the right of row α_i . Thus,

$$E(MS_a) = \sigma_e^2 + nD_qD_r\sigma_{\beta\gamma}^2 + nqD_r\sigma_{\alpha\gamma}^2 + nrD_q\sigma_\beta^2 + nqr\sigma_\alpha^2.$$

In words, the expected value of the mean square for the main effect of factor A is a weighted sum of a set of variances; the variances included in this set and the weights for each are determined by application of rule 6.

Since, in this design, factor B is nested under factor A , the main effect of factor B is denoted by the symbol $\beta_{j(i)}$. In this case the expected value of the mean square for the main effect of factor B is a weighted sum of the variances of terms containing both the subscripts i and j . These variances are

$$\sigma_e^2, \sigma_{\beta\gamma}^2, \text{ and } \sigma_\beta^2.$$

The weight for σ_β^2 is the product of all terms in row $\beta_{j(i)}$, the terms in columns i and j being omitted; this weight is nr . The weight for $\sigma_{\beta\gamma}^2$ is the product of all terms in row $\beta\gamma_{j(i)k}$, the terms in columns i and j being omitted. This weight is nD_r . The expected value for the mean square of the main effect of factor B (which is nested under factor A) is given at the right of the row $\beta_{j(i)}$.

The other terms under the heading $E(MS)$ in Table 5.14-1 were obtained by application of rules 9 and 10. For example, the variances that enter into the expected value of MS_{ac} are

$$\sigma_e^2, \sigma_{\beta\gamma}^2, \text{ and } \sigma_{\alpha\gamma}^2.$$

These are the variances corresponding to those effects which contain the subscripts i and k .

The expected values for the mean squares in Table 5.14-1 may be specialized to cover specific designs by evaluating the D 's. For example, if factors A and C are fixed and factor B is random, $D_p = 0$, $D_r = 0$, and $D_q = 1$. For this case,

$$E(MS_a) = \sigma_e^2 + nr\sigma_\beta^2 + nqr\sigma_\alpha^2.$$

This expected value was obtained from the corresponding expected value in Table 5.14-1 by evaluating the D_p , D_q , and D_r . By a similar procedure,

$$E(MS_b) = \sigma_e^2 + nr\sigma_\beta^2.$$

The expected values of mean squares for specialized designs may be obtained by evaluating the D 's prior to the application of the rules. For the

Table 5.14-2 Expected Value of Mean Squares for Partially Hierarchal Factorial Design
(Factors A and C fixed, factor B random)

Effect	i	j	k	m	$E(MS)$
α_i	0	q	r	n	$\sigma_e^2 + nr\sigma_\beta^2 + nqr\sigma_\alpha^2$
$\beta_{j(i)}$	1	1	r	n	$\sigma_e^2 + nr\sigma_\beta^2$
γ_k	p	q	0	n	$\sigma_e^2 + n\sigma_{\beta\gamma}^2 + np\sigma_\gamma^2$
$\alpha\gamma_{ik}$	0	q	0	n	$\sigma_e^2 + n\sigma_{\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
$\beta\gamma_{j(i)k}$	1	1	0	n	$\sigma_e^2 + n\sigma_{\beta\gamma}^2$
$\varepsilon_{m(ijk)}$	1	1	1	1	σ_e^2

special case under consideration, the expected values of the mean squares are derived in Table 5.14-2. The expected values of the mean squares shown on the right are identical to those obtained by specializing the $E(MS)$ given at the right of Table 5.14-1. For special cases of interest to the experimenter, the simplest method of deriving the expected values of the mean squares is that illustrated in Table 5.14-2.

5.15 Quasi F Ratios

In some cases the appropriate F ratio cannot be constructed by direct application of the rules based upon expected values of mean squares. For example, consider the case of a $p \times q \times r$ factorial experiment having n observations per cell. If all factors are random, the expected values of the mean squares are those given in Table 5.15-1. (These expected values correspond to model III in Table 5.9-2.)

In practice preliminary tests on the model (which are discussed in Sec. 5.16) are made on the higher-order interactions before proceeding with the tests on main effects. (In many cases which arise in practice, tests on main effects may be relatively meaningless when interactions are significantly

different from zero.) If none of the interaction terms may be dropped from the model, then no single mean square can serve as a denominator in a test on main effects due to factor A . The proper denominator for a test in this case should have the expected value

$$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2.$$

None of the individual mean squares in Table 5.15-1 has this expected value. However, by adding and subtracting certain of the mean squares one may

Table 5.15-1 Expected Values of Mean Squares
(Model III)

Source of variation	E(MS)
A	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
B	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + np\sigma_{\beta\gamma}^2 + nr\sigma_{\alpha\beta}^2 + npr\sigma_{\beta}^2$
C	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + np\sigma_{\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + npq\sigma_{\gamma}^2$
AB	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nr\sigma_{\alpha\beta}^2$
AC	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
BC	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + np\sigma_{\beta\gamma}^2$
ABC	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$
Experimental error	σ_e^2

obtain a composite mean square which has the required expected value. One such composite, assuming each of the mean squares is independent of the others, may be constructed as follows:

$$E(MS_{ac}) = \sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$$

$$E(MS_{ab}) = \sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nr\sigma_{\alpha\beta}^2$$

$$-E(MS_{abc}) = -\sigma_e^2 - n\sigma_{\alpha\beta\gamma}^2$$

$$\overline{E(MS_{ac} + MS_{ab} - MS_{abc})} = \sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2$$

A quasi F ratio, which has the proper structural requirements in terms of expected values of mean squares, for a test on the main effect of factor A is

$$F' = \frac{MS_a}{MS_{ac} + MS_{ab} - MS_{abc}}.$$

The symbol F' is used for this ratio rather than the symbol F . Since the denominator is a composite of different sources of variation, the sampling distribution of the F' ratio is not the usual F distribution, although the latter distribution may be used as an approximation. The denominator of this F' ratio calls for subtracting a mean square. This could lead to the possibility of obtaining a negative denominator. According to the population model, it is not possible to have a negative denominator in terms of parameters.

However, in terms of the estimates of these parameters it is possible to have a negative denominator.

The following quasi F ratio avoids the possibility of a negative denominator and still satisfies the structural requirements for a test on the main effect of factor A :

$$F'' = \frac{MS_a + MS_{abc}}{MS_{ac} + MS_{ab}}$$

In terms of expected values of the mean squares, this ratio has the form

$$F'' = \frac{2\sigma_\epsilon^2 + 2n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2 + nqr\sigma_\alpha^2}{2\sigma_\epsilon^2 + 2n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2}$$

Under the hypothesis that $\sigma_\alpha^2 = 0$, $E(F'') = 1$. Similarly, the following F'' ratio has the structural requirements in terms of the expected values of mean squares for a test of the hypothesis that $\sigma_\beta^2 = 0$:

$$F'' = \frac{MS_b + MS_{abc}}{MS_{ab} + MS_{bc}}$$

Although these F'' ratios satisfy the structural requirements in terms of expected values, the sampling distributions of these F'' ratios can only be roughly approximated by the usual F distributions, provided that special degrees of freedom are used for numerator and denominator. Suppose that the F'' ratio has the following general form,

$$F'' = \frac{u + v}{w + x},$$

where u, v, w , and x are appropriate mean squares. Let respective degrees of freedom for these mean squares be f_u, f_v, f_w , and f_x . Then the degrees of freedom for the numerator are approximated by the nearest integral value to

$$\frac{(u + v)^2}{(u^2/f_u) + (v^2/f_v)}$$

The degrees of freedom for the denominator are approximated by

$$\frac{(w + x)^2}{(w^2/f_w) + (x^2/f_x)}$$

This approximation to the F distribution and the associated degrees of freedom are those suggested by Satterthwaite (1946).

The F' ratio has the following general form:

$$F' = \frac{u}{w + x - v}$$

If the F distribution is used to approximate the sampling distribution of the F' statistic, the degrees of freedom for the denominator are

$$\frac{(w + x - v)^2}{(w^2/f_w) + (x^2/f_x) + (v^2/f_v)}.$$

5.16 Preliminary Tests on the Model and Pooling Procedures

In deriving the expected values for the mean squares, extensive use is made of a model which is assumed to be appropriate for an experiment. The model indicates the relevant sources of variability. A question might be raised about why certain terms appear in the model and why other possible terms are omitted. If in fact there is no interaction effect of a given kind in the population of interest, why should such an interaction term be included in the model? Including such an interaction term in the model can potentially affect the expected values of several mean squares. The latter in turn determine the structure of F ratios.

Decisions about what terms should appear in the model and what terms should be omitted are generally based upon experience in an experimental area and knowledge about what are reasonable expectations with respect to underlying sources of variation—in short, subject-matter information. All sources of variation not specifically included in the model are in reality classified as part of the experimental error. In most cases the latter variation is the residual variation after all controlled sources have been estimated. Previous experimentation in an area may indicate that no interaction between two factors is to be expected; hence in designing a new experiment in this area such an interaction term may be omitted from the model. However, any variation due to this interaction, if it exists, is automatically included as part of the experimental error or automatically confounds other estimates, depending upon the experimental design.

Lacking knowledge about interaction effects from past experimentation, one might ask whether or not data obtained in a given factorial experiment could be used as a basis for revising an initial model. The specification of the parameters in the initial model could be considered incomplete or left open to more complete specification. Tests designed to revise or complete the specification of parameters to be included in the model are called *preliminary* tests on the model. Such tests are particularly appropriate when one is dealing with experiments in which interactions between fixed and random factors or interactions between random factors are potentially in the model. Such terms may turn out to be denominators for F ratios. If such terms have a relatively small number of degrees of freedom, corresponding F ratios will have very low power. Since the mean squares for interactions between two or more fixed factors can never form the denominator in an F ratio, preliminary tests on such interactions are not generally required.

The procedure of making preliminary tests on higher-order interactions before proceeding with tests of lower-order interactions and main effects may be regarded as a multistage decision rule. Depending upon the outcome of a sequence of tests, the parameters in the model become more completely specified; in turn the expected values for the mean squares are revised sequentially.

If a preliminary test does not reject the hypothesis that the variance due to an interaction effect is zero, one proceeds as if this variance were actually zero and drops the corresponding term from the model. The expected values of the mean squares are then revised in accordance with the new model, and additional preliminary tests, if necessary, are made. Care must

Table 5.16-1 Expected Values of Mean Squares for Model in (1)
(A fixed, B and C random)

Source of variation	df	E(MS)
A	$p - 1$	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
B	$q - 1$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2 + npr\sigma_{\beta}^2$
C	$r - 1$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2 + npq\sigma_{\gamma}^2$
AB	$(p - 1)(q - 1)$	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nr\sigma_{\alpha\beta}^2$
AC	$(p - 1)(r - 1)$	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
BC	$(q - 1)(r - 1)$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2$
ABC	$(p - 1)(q - 1)(r - 1)$	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$
Within cell	$pqr(n - 1)$	σ_e^2

be taken in such tests to avoid type 2 error, i.e., accepting the hypothesis of zero interaction when it should be rejected. Type 2 error can be kept numerically small by making preliminary tests at a numerically high type 1 error, that is, $\alpha = .20$ or $.30$.

The sequence in which preliminary tests on the model are made will be illustrated for the case of a $p \times q \times r$ factorial experiment in which factor A is considered fixed and factors B and C are considered random. The initial model for this experiment is the following:

$$(1) \quad X_{ijklm} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \varepsilon_{m(ijk)}.$$

The question of whether or not all the interactions between random factors and between fixed and random factors should be included in the model is left unanswered for the time being. Assuming the complete model, the expected values of the mean squares are those given in Table 5.16-1. Suppose that the experimental work is completed and the analysis-of-variance table prepared. Tests of hypotheses depend upon the model and associated expected values. The model as given in (1) is tentative and subject to change. Associated with this model are the expected values given in Table 5.16-1. Inspection of this model indicates that the appropriate

denominator for tests on the hypotheses $\sigma_{\alpha\beta\gamma}^2 = 0$ and $\sigma_{\beta\gamma}^2 = 0$ is $MS_{w. cell}$. Suppose both of these tests are made at the 25 per cent level of significance. Suppose that these tests do not reject the respective hypotheses and that a priori information indicates no good basis for expecting that such interactions exist. On these grounds $\alpha\beta\gamma_{ijk}$ and $\beta\gamma_{jk}$ are now dropped from the model.

The revised model now has the form

$$(2) \quad X_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \varepsilon_{ijkm}.$$

The term ε now includes the interaction terms which were dropped from the model in (1). The revised expected values of the mean squares are obtained

Table 5.16-2 Expected Values of Mean Squares for Model in (2)

Source of variation	df	E(MS)
<i>A</i>	$p - 1$	$\sigma_e^2 + nq\sigma_{\alpha\gamma}^2 + nr\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
<i>B</i>	$q - 1$	$\sigma_e^2 + npr\sigma_{\beta}^2$
<i>C</i>	$r - 1$	$\sigma_e^2 + npq\sigma_{\gamma}^2$
<i>AB</i>	$(p - 1)(q - 1)$	$\sigma_e^2 + nr\sigma_{\alpha\beta}^2$
<i>AC</i>	$(p - 1)(r - 1)$	$\sigma_e^2 + nq\sigma_{\alpha\gamma}^2$
<i>BC</i>		
<i>ABC</i>		
Within cell } residual	$npqr - pq - pr + p$	σ_e^2

by dropping $\sigma_{\alpha\beta\gamma}^2$ and $\sigma_{\beta\gamma}^2$ from the terms in Table 5.16-1. When this is done, MS_{bc} , MS_{abc} , and $MS_{w. cell}$ are all estimates of σ_e^2 . These three mean squares may be pooled to provide a single estimate of σ_e^2 as shown in Table 5.16-2. The pooled estimate is

$$MS_{res} = \frac{SS_{bc} + SS_{abc} + SS_{w. cell}}{(q - 1)(r - 1) + (p - 1)(q - 1)(r - 1) + pqr(n - 1)}.$$

The degrees of freedom for the denominator are the sum of the degrees of freedom for the sources of variation which are pooled. This sum is equal to $npqr - pq - pr + p$.

Inspection of Table 5.16-2 indicates that the hypotheses $\sigma_{\alpha\gamma}^2 = 0$ and $\sigma_{\alpha\beta}^2 = 0$ may both be tested with MS_{res} as a denominator. Suppose that these tests are made at the 25 per cent level of significance. Suppose that the outcome of these tests does not reject the hypothesis that $\sigma_{\alpha\gamma}^2 = 0$ but that the hypothesis that $\sigma_{\alpha\beta}^2 = 0$ is rejected. The revised model now has the form

$$(3) \quad X_{ijkm} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \varepsilon_{ijkm}.$$

The experimental error in (3) includes variation due to $\alpha\gamma$ as well as the interaction terms included in model (2). The expected values associated

with (3) are given in Table 5.16-3. These expected values may be taken as those appropriate for final tests.

If there is a priori evidence to indicate interaction between factors, preliminary tests on such interactions should in general be avoided and tests

Table 5.16-3 Expected Values of Mean Squares for Model in (3)

Source of variation	df	E(MS)
<i>A</i>	$p - 1$	$\sigma_e^2 + nro_{\alpha\beta}^2 + nqr\sigma_\alpha^2$
<i>B</i>	$q - 1$	$\sigma_e^2 + npro_{\beta}^2$
<i>C</i>	$r - 1$	$\sigma_e^2 + npq\sigma_\gamma^2$
<i>AB</i>	$(p - 1)(q - 1)$	$\sigma_e^2 + nro_{\alpha\beta}^2$
Residual (pooled <i>AC</i> , <i>BC</i> , <i>ABC</i> , and within cell)	$npqr - pq - r + 1$	σ_e^2

should be made in accordance with the original formulation. When preliminary tests are made, only interactions in which random factors appear are considered in such tests. (Only interactions with random factors can potentially form the denominator of an F ratio.) If an adequate number of degrees of freedom (say, 20 or more) is available for the denominator of F ratios constructed in terms of the original model, preliminary tests should also be avoided. However, if the denominator of an F ratio constructed in accordance with the original model has relatively few degrees of freedom (say, less than 10), in the absence of a priori knowledge preliminary tests are in order.

A numerical example will be used to illustrate the pooling procedures associated with Tables 5.16-1, 5.16-2, and 5.16-3. Part i of Table 5.16-4 represents the analysis of variance for a $4 \times 2 \times 3$ factorial experiment having six observations per cell. Assume factor *A* fixed and factors *B* and *C* random. If the model in (1) is assumed, the expected values of the mean squares given in Table 5.16-1 are appropriate. Preliminary tests on $\sigma_{\alpha\beta\gamma}^2$ and $\sigma_{\beta\gamma}^2$ are made in part i. These tests indicate that the hypotheses that $\sigma_{\alpha\beta\gamma}^2 = 0$ and $\sigma_{\beta\gamma}^2 = 0$ cannot be rejected at the 25 per cent level of significance. Hence the corresponding terms are dropped from the model in (1).

The residual term in part ii is obtained as follows:

Source	SS	df	MS
<i>BC</i>	18.00	2	
<i>ABC</i>	72.00	6	
Within cell	240.00	24	
Residual (ii)	330.00	32	10.31

The variation due to residual ii corresponds to variation due to experimental error in the model given in (2). The expected values for the mean

Table 5.16-4 Numerical Example of Pooling Procedures

 $(n = 2, p = 4, q = 2, r = 3)$ First-stage preliminary tests: $H_1: \sigma_{\alpha\beta\gamma}^2 = 0$; $H_1: \sigma_{\beta\gamma}^2 = 0$

Source	SS	df	MS	F
A	120.00	3	40.00	
B	60.00	1	60.00	
C	40.00	2	20.00	
AB	96.00	3	32.00	
(i) AC	72.00	6	12.00	
BC	18.00	2	9.00	0.90 $F_{.75}(2,24) = 1.47$
ABC	72.00	6	12.00	1.20 $F_{.75}(6,24) = 1.41$
Within cell	240.00	24	10.00	

Second-stage preliminary tests: $H_1: \sigma_{\alpha\beta}^2 = 0$; $H_1: \sigma_{\alpha\gamma}^2 = 0$

Source	SS	df	MS	F
A	120.00	3	40.00	
B	60.00	1	60.00	
C	40.00	2	20.00	
(ii) AB	96.00	3	32.00	3.10 $F_{.75}(3,32) = 1.44$
AC	72.00	6	12.00	1.16 $F_{.75}(6,32) = 1.39$
Residual (ii)	330.00	32	10.31	

Analysis for final tests:

Source	SS	df	MS	F
A	120.00	3	40.00	1.25 $F_{.95}(3,3) = 9.28$
B	60.00	1	60.00	5.67 $F_{.95}(1,38) = 4.10$
(iii) C	40.00	2	20.00	1.89 $F_{.95}(2,38) = 3.25$
AB	96.00	3	32.00	3.02 $F_{.95}(3,38) = 2.85$
Residual (iii)	402.00	38	10.58	

squares in part ii are those given in Table 5.16-2. Second-stage tests are made in accordance with the latter expected values. As a result of the tests in part ii, the variation associated with the AC interaction is pooled with the experimental error, but the variation due to the AB interaction is not pooled.

The residual term in part iii is obtained as follows:

Source	SS	df	MS
AC	72.00	6	
Residual (ii)	330.00	32	
Residual (iii)	402.00	38	10.58

Statistical tests in part iii are based upon the expected values in Table 5.16-3. The denominator for the F ratio in the test on the main effect of factor A is MS_{ab} ; all other F ratios have MS_{res} as a denominator. As distinguished from parts i and ii, the tests in part iii are final tests rather than preliminary tests; hence the difference in the level of significance.

The sampling distributions of the statistics used in tests made following preliminary tests are actually different from the sampling distributions associated with tests which are not preceded by preliminary tests. What is really required in the second stage of a sequential decision procedure is the sampling distribution of the statistic in question under the condition that specified decisions have been made in the first stage. Using sampling distributions which do not have such conditions attached generally introduces a slight bias into the testing procedure. Specified percentile points on the unconditional sampling distribution are probably slightly lower than corresponding points on the conditional sampling distribution. That is, a statistic which falls at the 95th percentile point when referred to the unconditional distribution may fall at only the 92d percentile point when referred to the conditional distribution. Hence use of the unconditional distributions for sequential tests probably gives tests which have a slight positive bias; i.e., the type 1 error is slightly larger than the specified level of significance.

By way of summary, it should be noted that there is no widespread agreement among statisticians on the wisdom of the pooling procedures which have been discussed in this section. Those statisticians who adhere to the "never pool" rule demand a completely specified model prior to the analysis of the experimental data. This position has much to recommend it. The inferences obtained from adopting this point of view will be based upon exact sampling distributions, provided that the model that has been specified is appropriate for the experiment.

Using data from the experiment to revise the model introduces contingencies which are difficult to evaluate statistically. However, working from a revised model which more adequately fits the data may potentially provide more powerful tests than those obtained from the "never pool" rule. Admittedly the change in power cannot be evaluated with precision. The conservative attitude toward pooling adopted in this section attempts to take middle ground: one departs from the initial model only if the experimental data strongly suggest that the initial model is not appropriate. The position taken by the author in this section is quite close to that adopted by Green and Tukey (1960). It is also in line with the point of view developed by Bozovich et al. (1956).

5.17 Individual Comparisons

Procedures discussed in Chap. 3 for making individual and multiple comparisons between means can be extended rather directly to factorial

experiments. A significant over-all F test on a main effect, for example, indicates that one or more of a multitude of possible comparisons is significant. The specific comparisons which are built into the design or suggested by the theoretical basis for the experiment can and should be made individually, regardless of the outcome of the corresponding over-all F test. Seldom, if ever, should a posteriori comparisons be made when the over-all F is nonsignificant. (In this context an a posteriori comparison is one suggested by deliberate inspection of the data.) Should such comparisons be made, statistically significant outcomes should be interpreted with extreme caution. The experimenter should not hesitate to describe fully all aspects of his experimental results.

The procedure for making individual comparisons will be illustrated for the case of a $p \times q$ factorial experiment having n observations per cell. To test the hypothesis that $\alpha_i = \alpha_{i'}$ against the two-tailed alternative hypothesis $\alpha_i \neq \alpha_{i'}$, one may use the test statistic

$$F = \frac{(\bar{A}_i - \bar{A}_{i'})^2}{MS_{\bar{A}_i - \bar{A}_{i'}}}.$$

The best estimate of $MS_{\bar{A}_i - \bar{A}_{i'}}$ depends upon whether factor B is fixed or random. For the case in which factor B is fixed,

$$MS_{\bar{A}_i - \bar{A}_{i'}} = \frac{2MS_{w, \text{cell}}}{nq}.$$

For the case in which factor B is random and $\sigma_{\alpha\beta}^2 \neq 0$,

$$MS_{\bar{A}_i - \bar{A}_{i'}} = \frac{2MS_{ab}}{nq}.$$

(When the AB interaction is significant, one is generally interested only in the simple main effects. Hence tests on the over-all main effects in the presence of significant interaction seldom are made in practice.) For the first case, the degrees of freedom for $MS_{\bar{A}_i - \bar{A}_{i'}}$ are $pq(n-1)$; for the second case, the degrees of freedom are $(p-1)(q-1)$. In either case

$$t = \sqrt{F} = \frac{\bar{A}_i - \bar{A}_{i'}}{\sqrt{MS_{\bar{A}_i - \bar{A}_{i'}}}}.$$

Either one-tailed or two-tailed tests may be made by using the t statistic. An equivalent, but computationally simpler form of the F statistic for the case in which factor B is fixed, in terms of totals rather than means, is

$$F = \frac{(A_i - A_{i'})^2}{2nqMS_{w, \text{cell}}}.$$

For the case in which factor B is random,

$$F = \frac{(A_i - A_{i'})^2}{2nqMS_{ab}}.$$

Regardless of whether factors A or B are fixed or random, a test on the difference between cell means has the form

$$F = \frac{(AB_{ij} - AB_{km})^2}{2nMS_{w. cell}}.$$

In making the tests indicated in this section, the denominators use data from all cells in the experiment, not just those from which the means have been computed. If the experimental error is homogeneous, use of information from all cells to estimate experimental error is justified. If, however, there is a sufficient number of degrees of freedom (say, over 30) for the estimation of experimental error from only those cells which are used in the estimation of the means being compared, this latter estimate is to be preferred to the pooled estimate from all cells.

A test on the difference between two means is a special case of a comparison among several means. For $p \times q$ factorial experiments having n observations in each cell, the mean square for a comparison among the main effects of factor A has the general form

$$\frac{(c_1A_1 + c_2A_2 + \cdots + c_pA_p)^2}{nq\Sigma c_i^2},$$

where $\Sigma c_i = 0$. The case in which $c_1 = 1$, $c_2 = -1$, and all other c 's are zero defines the mean square

$$\frac{(A_1 - A_2)^2}{2nq}.$$

The case in which $c_1 = 1$, $c_2 = 1$, and $c_3 = -2$ defines the mean square

$$\frac{(A_1 + A_2 - 2A_3)^2}{6nq}.$$

The general mean square corresponding to a comparison among the main effects of factor B has the form

$$\frac{(c_1B_1 + c_2B_2 + \cdots + c_qB_q)^2}{np\Sigma c_j^2},$$

where $\Sigma c_j = 0$. Similarly, the mean square for a comparison among the simple effects of factor A for level b_j has the form

$$\frac{(c_1AB_{1j} + c_2AB_{2j} + \cdots + c_pAB_{pj})^2}{n\Sigma c_i^2}.$$

For the case in which A and B are fixed factors, comparisons planned before the data were obtained (and assuming that the number of such comparisons is small relative to the total number possible) can be made by use of the following statistic:

$$F = \frac{MS_{\text{comparison}}}{MS_{w. cell}}.$$

Under the assumption that the hypothesis being tested is true, this statistic has a sampling distribution which is given by an F distribution. The critical value for this test is $F_{1-\alpha}[1, pq(n-1)]$. In cases where the comparisons are large in number or of the a posteriori type, the appropriate critical value, as suggested by Scheffé (1960), for tests on main effects due to factor A is

$$(p-1)F_{1-\alpha}[(p-1), pq(n-1)].$$

An analogous critical value for comparisons among the main effects of factor B is

$$(q-1)F_{1-\alpha}[(q-1), pq(n-1)].$$

The critical value for comparisons of the a posteriori type among cell means is

$$(pq-1)F_{1-\alpha}[(pq-1), pq(n-1)].$$

To illustrate the magnitude of the difference between the two types of critical values, consider the case in which $p = 10$ and $pq(n-1) = 40$. For $\alpha = .01$,

$$F_{.99}(1,40) = 5.42$$

is the critical value for an a priori type of comparison. For an a posterior type

$$9F_{.99}(9,40) = 9(2.45) = 22.05.$$

The difference between the two critical values is quite marked—but then the difference in the underlying logic is also quite marked. Whenever a relatively large number of tests of significance are to be made, or whenever comparisons suggested by the data are made, the usual sampling distributions (that is, t or F) associated with tests of significance no longer apply. The critical value for the a posteriori comparison is much larger because the sampling distribution appropriate for this type of comparison must take into account sources of variation which are not relevant to the a priori comparison.

In making comparisons between all possible pairs of ordered means within a logical grouping, the Tukey or the Newman-Keuls procedures may be adapted for use. For example, all pairs of means within a single row or column of a $p \times q$ factorial experiment may be compared by means of the Tukey procedure. If the number of degrees of freedom for estimating the standard error is sufficiently large (say, over 20), a single row or a single column of the $p \times q$ factorial experiment may be considered as a single-factor experiment and the methods in Chap. 3 may be applied directly to this part of the factorial experiment. Alternatively, if the assumption of homogeneity of within-cell variance is met, an estimate of the standard error of a cell mean is given by

$$s_{\overline{AB}} = \sqrt{\frac{MS_{w, cell}}{n}}.$$

The degrees of freedom for this estimate are $pq(n-1)$.

If all possible pairs of means corresponding to the main effects of factor A are to be compared, and if factor B is fixed, then

$$s_{\bar{A}} = \sqrt{\frac{MS_{w, \text{cell}}}{nq}}.$$

Numerical applications of comparisons of this kind are given in Sec. 6.2.

5.18 Partition of Main Effects and Interaction into Trend Components

In some cases arising in practice it is desirable to divide main effects as well as interactions into components associated with functional forms assumed to account for trends in the criterion responses. As an example, consider a 3×4 factorial experiment in which the levels of both factors may be regarded as steps along an essentially underlying continuum. The magnitudes of the criterion scores within each of the cells may be considered to define a response surface. It is frequently of interest to explore regions on this response surface, particularly when one is seeking an optimum combination of treatments.

Main effects and interactions in a 3×4 factorial experiment may be subdivided as indicated below:

Source		df
A		2
Linear	1	
Quadratic	1	
B		3
Linear	1	
Quadratic	1	
Cubic	1	
AB		6
Linear \times linear	1	Quadratic \times linear 1
Linear \times quadratic	1	Quadratic \times quadratic 1
Linear \times cubic	1	Quadratic \times cubic 1

The actual equation (or surface) which will describe each of the cell means exactly has the form

$$\bar{X} = c_0 + c_1A + c_2B + c_3A^2 + c_4B^2 + c_5B^3 + c_6AB + c_7AB^2 + c_8AB^3 \\ + c_9A^2B + c_{10}A^2B^2 + c_{11}A^2B^3.$$

Each term, except c_0 , in the equation corresponds to a term in the partition of the 11 degrees of freedom into trend components. For example, AB^2 corresponds to the linear \times quadratic interaction; A^2B^3 corresponds to the quadratic \times cubic interaction.

The components of variation defining the A main effects have the general form

$$A \text{ (component)} = \frac{(\sum c_i A_i)^2}{nq \sum c_i^2},$$

where $\Sigma c_i = 0$. The components of the main effect of factor B have the following general form,

$$B(\text{component}) = \frac{(\Sigma c_j B_j)^2}{np \Sigma c_j^2},$$

where $\Sigma c_j = 0$. The form of the components of the AB interaction is

$$AB(\text{component}) = \frac{(\Sigma c_i c_j AB_{ij})^2}{n \Sigma (c_i c_j)^2},$$

where $\Sigma c_i = 0$ and $\Sigma c_j = 0$. The computational work is simplified if the levels of factors A and B are equally spaced along the underlying continuum; in this case coefficients of orthogonal polynomials define the c_i and c_j . In case of unequal spacing, these coefficients may be modified by using methods described by Robson (1959).

The components of the interaction obtained in this way describe features of the response surface. More extensive methods for exploring response surfaces are available. A survey of some of these methods is given in Cochran and Cox (1957, chap. 8A).

If only one of the factors, say, factor B , is continuous, then the following analysis is possible:

Source	df
A	2
B	3
Linear	1
Quadratic	1
Cubic	1
AB	6
Difference in lin trend	2
Difference in quad trend	2
Difference in cubic trend	2

This method of partitioning the interaction is particularly appropriate in attempting to interpret differences in shapes of profiles. Computational details for this kind of partition are given in Secs. 6.9 and 7.6. In Sec. 7.6, computational steps are given in terms of a design calling for repeated measures. If the design does not have repeated measures, computational procedures for treatment effects remain the same but no partition of the error variation is required.

For the case of a 2×2 factorial experiment in which both factors represent what can be considered an underlying continuum, an exact fit to the surface defined by the four cell means is given by the polynomial

$$\bar{X} = c_0 + c_1 A + c_2 B + c_3 AB.$$

For the case of a 2×3 factorial experiment, an exact fit to the surface defined by the six cell means is given by the polynomial

$$\bar{X} = c_0 + c_1A + c_2B + c_3B^2 + c_4AB + c_5AB^2.$$

The term AB corresponds to the linear \times linear interaction; the term AB^2 corresponds to the linear \times quadratic interaction.

Surfaces defined by polynomials which have as many terms as there are cell means are of no predictive value. Zero degrees of freedom remain for estimating standard errors; in essence such standard errors are infinitely large. The smaller the number of parameters required in defining the form of the appropriate response surface, the greater, potentially, the predictive value of the resulting equation.

5.19 Replicated Experiments

A replication of an experiment is an independent repetition under as nearly identical conditions as the nature of the experimental material will permit. *Independent* in this context implies that the experimental units in the repetitions are independent samples from the population being studied. That is, if the elements are people, in a replication of an experiment an independent sample of people is used in the replication. Inferences made from replicated experiments are with respect to the outcomes of a series of replications of the experiment. In a sense that will be indicated, inferences from replicated experiments have a broader scope than do inferences from nonreplicated experiments.

An experiment in which there are n observations per cell is to be distinguished from an experiment having n replications with one observation per cell. The total number of observations per treatment is the same, but the manner in which the two types of experiments are conducted differs. Consequently the relevant sources of variation will differ; hence a change in the model for the experiment is needed. Associated with the latter change is a different set of expected values for the mean squares.

In conducting an experiment having n observations per cell, all observations in a single cell are generally made within the same approximate time interval or under experimental conditions that can be considered to differ only as a function of experimental error. In a $p \times q$ factorial experiment having n observations per cell (no repeated measures) the total variation may be subdivided as follows.

Source	df
Between cells	$pq - 1$
Within cells	$pq(n - 1)$

In this partition it is assumed that differences between observations within the same cell are attributable solely to experimental error. On the other hand, differences among cells are attributed to treatment effects and experimental error. Should there be systematic sources affecting between-cell

variation, other than treatment effects and experimental error, such variation is completely or partially confounded with the treatment effects.

The purpose of a replicated experiment, in contrast to an experiment having n observations per cell, is to permit the experimenter to maintain more uniform conditions within each cell of the experiment, as well as to eliminate possible irrelevant sources of variation between cells. For example, the replications may be repetitions of the experiment at different times or at different places. In this case, sources of variation which are functions of time or place are eliminated from both between- and within-cell variation.

Instead of having n replications with one observation per cell, one may have two replications in which there are $n/2$ observations per cell, one may have three replications with $n/3$ observations per cell, etc. The number of observations per cell for any single replication should be the maximum that will permit uniform conditions within all cells of the experiment and, at the same time, reduce between-cell sources of variation which are not directly related to the treatment effects.

As an example of a replicated experiment, suppose that it is desired to have 15 observations within each cell of a 2×3 factorial experiment. Suppose that the experimental conditions are such that only 30 observations can be made within a given time period; suppose further that there are, potentially, differences in variation associated with the time periods. To eliminate sources of variation directly related to the time dimension, the experiment may be set up in three replications as follows (cell entries indicate the number of observations):

Replication 1				Replication 2				Replication 3			
	b_1	b_2	b_3		b_1	b_2	b_3		b_1	b_2	b_3
a_1	5	5	5	a_1	5	5	5	a_1	5	5	5
a_2	5	5	5	a_2	5	5	5	a_2	5	5	5

If the replications are disregarded, there are 15 observations under each treatment combination. Within each replication, conditions may be relatively uniform with respect to uncontrolled sources of variation; between replications, conditions may not be uniform. In essence, a replicated experiment adds another dimension or factor to the experiment—a replication factor. The latter is a random factor.

In Table 5.19-1 the analysis of a $p \times q$ factorial experiment having n observations per cell is contrasted with a $p \times q$ factorial experiment in which there are r replications of a $p \times q$ factorial experiment having n/r observations per cell. In both designs there are n observations under treatment combination and a total of $npq - 1$ degrees of freedom. The degrees of freedom in braces in part i are often pooled in a replicated experiment. Part ii gives an alternative partition of the variation. In the non-replicated experiment the between-cell variation defines the estimates of

variation due to treatment effects. On the other hand, in the replicated experiment part of the between-cell variation is due to replication effects as well as interactions with replications. The within-cell variations in both designs is considered to be due to experimental error. Since observations

Table 5.19-1 Comparison of Analysis for Replicated and Nonreplicated Factorial Experiments

$p \times q$ factorial experiment with n observations per cell		r replications of $p \times q$ factorial experiment with n/r observations per cell	
	df		df
A	$p - 1$	A	$p - 1$
B	$q - 1$	B	$q - 1$
AB	$(p - 1)(q - 1)$	AB	$(p - 1)(q - 1)$
(i)		Reps	$r - 1$
Within cell	$pq(n - 1)$	$A \times \text{rep}$	$(p - 1)(r - 1)$
		$B \times \text{rep}$	$(q - 1)(r - 1)$
		$AB \times \text{rep}$	$(p - 1)(q - 1)(r - 1)$
Total	$npq - 1$	Within cell	$pqr[(n/r) - 1]$
		Total	$npq - 1$
Between cells	$pq - 1$	Between cells	$pqr - 1$
(ii)		Treatments	$pq - 1$
Within cell	$pq(n - 1)$	Reps	$r - 1$
		Treat \times rep	$(pq - 1)(r - 1)$
		Within cell	$pqr[(n/r) - 1]$

in the replicated experiment are made under somewhat more controlled conditions, the within-cell variation in a replicated experiment is potentially smaller than the corresponding variation in a nonreplicated experiment.

The partition of the total variation given in part ii may be illustrated numerically. Consider a 3×4 factorial experiment in which there are to be 10 observations under each treatment combination. The design calling for 10 observations per cell is contrasted with the design calling for five replications with 2 observations per cell in the following partition:

10 observations per cell	5 reps, 2 observations per cell
Between cells	11
Within cell	108
	Between cells
	Treatments
	Reps
	Treat \times rep
	Within cell

In a replicated design of this kind interactions with replications are often considered to be part of the experimental error. However, preliminary

tests on the model may be used to check upon whether or not such pooling is justified.

There are both advantages and disadvantages to replicated experiments when contrasted with those in which there are n observations per cell. The advantages of being able to eliminate sources of variation associated with replications depend upon the magnitude of the latter relative to treatment effects and effects due to experimental error. One possible disadvantage of a replicated design might arise in situations requiring precise settings of experimental equipment. Considerable time and effort may be lost in resetting equipment for each of the replications, rather than making all observations under a single set of experimental conditions before moving to the next experimental condition. However, the inferences which can be drawn from the replicated experiment are stronger when possible variations in the resettings are considered as a relevant source of variation in the conduct of the experiment.

5.20 The Case $n = 1$ and a Test for Nonadditivity

If there is only one observation in each cell of a $p \times q$ factorial experiment, there can be no within cell variation and hence no direct estimate of experimental error. Among other models, the following two may be postulated to underlie the observed data:

- (i)
$$X_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}.$$
- (ii)
$$X_{ij} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}.$$

In (i) no interaction effect is postulated, hence all sources of variation other than main effects are considered to be part of the experimental error. In (ii) an interaction term is postulated. From some points of view the interaction term may be considered as a measure of nonadditivity of the main effects. In this context, (i) will be considered the additive model, whereas (ii) will be considered the nonadditive model.

The choice of the scale of measurement for the basic criterion measurement will to a large extent determine whether (i) or (ii) is the more appropriate model. The degree of appropriateness is gauged in terms of the degree of heterogeneity of the experimental error under model (i). Subject-matter knowledge and experience gained from past experimentation about the functional form of the underlying sources of variation are the best guides for deciding between models (i) and (ii). To supplement these sources of information, the experimenter may want to use information provided by preliminary tests before specifying the model. So far in this chapter such tests have been considered only for the case in which direct estimates of experimental error were available from the experimental data. Tukey (1949) has developed a test applicable to the case in which there is a single observation per cell. This test is discussed in detail by Scheffé (1960, pp. 130–134).

Tukey's test is called a test for *nonadditivity*. Its purpose is to help in the decision between models (i) and (ii). This test has also been used to choose between alternative possibilities for the scale for measurement for the criterion, the decision being made in favor of the scale for which (i) is the more appropriate model.

The principle underlying Tukey's test for nonadditivity is this: Assuming model (i) to hold, the variation due to sources other than main effects is divided into two parts. One part corresponds to the equivalent to the linear \times linear component of the AB interaction. This part measures what is called nonadditivity. The second part, or balance, is what is left. Should the component measuring nonadditivity be significantly larger than the balance, a decision is made against model (i). The component associated with nonadditivity is defined to be

$$MS_{\text{nonadd}} = \frac{(\sum \sum c_i c_j AB_{ij})^2}{\sum \sum (c_i c_j)^2},$$

where $c_i = (\bar{A}_i - \bar{G})$ and $c_j = (\bar{B}_j - \bar{G})$. An equivalent form for this last expression is

$$SS_{\text{nonadd}} = \frac{[\sum (\bar{A}_i - \bar{G})(\bar{B}_j - \bar{G})X_{ij}]^2}{SS_a SS_b / pq}.$$

Because of the way in which the c 's are defined, it will be found that

$$\sum \sum (c_i c_j)^2 = (\sum c_i^2)(\sum c_j^2).$$

In the application of Tukey's test for nonadditivity to the case of a $p \times q$ factorial experiment having one observation per cell, the analysis of variance takes the following form:

Source of variation	SS	df	MS
<i>A</i>	SS_a	$p - 1$	MS_a
<i>B</i>	SS_b	$q - 1$	MS_b
Residual	SS_{res}	$(p - 1)(q - 1)$	
Nonadditivity	SS_{nonadd}	1	MS_{nonadd}
Balance	SS_{bal}	$(p - 1)(q - 1) - 1$	MS_{bal}

The test for nonadditivity is given by

$$F = \frac{MS_{\text{nonadd}}}{MS_{\text{bal}}}.$$

When this F ratio exceeds the critical value defined by the level of significance of the test, the hypothesis that model (i) is appropriate is rejected. Tukey's test for nonadditivity is sensitive to only one source of nonadditivity—that associated with what is equivalent to a linear \times linear component of an interaction. In working with this component there is an implicit assumption that, the larger the main effects of the individual levels in a treatment

combination, the larger the potential interaction effect for the treatment combination, if this does exist. This assumption appears reasonable in some cases. In other cases it might be that the equivalent of the linear \times quadratic or the quadratic \times quadratic component could more appropriately be used as a measure of nonadditivity. A numerical example is given in Sec. 6.8.

The principles underlying Tukey's test can be extended to three-factor as well as higher-order factorial experiments. For three-factor experiments the measure of nonadditivity is given by

$$SS_{\text{nonadd}} = \frac{[\sum \sum \sum (c_i c_j c_k X_{ijk})]^2}{\sum \sum \sum (c_i c_j c_k)^2},$$

where $c_i = \bar{A}_i - \bar{G}$, $c_j = \bar{B}_j - \bar{G}$, and $c_k = \bar{C}_k - \bar{G}$. An equivalent form of this last expression is

$$SS_{\text{nonadd}} = \frac{(pqr)^2 [\sum \sum \sum (c_i c_j c_k X_{ijk})]^2}{(SS_a)(SS_b)(SS_c)}.$$

Because of the way in which the c 's are defined, it will be found that

$$\sum \sum \sum (c_i c_j c_k)^2 = (\sum c_i^2)(\sum c_j^2)(\sum c_k^2).$$

5.21 The Choice of a Scale of Measurement and Transformations

In the analysis of variance of a factorial experiment, the total variation of the criterion variable is subdivided into nonoverlapping parts which are attributable to main effects, interactions, and experimental error. The relative magnitude of each of the corresponding variances depends upon the scale of measurement as well as the spacing of the levels of the factors used in the experiment. In cases where alternative choices of a scale of measurement appear equally justifiable on the basis of past experience and theory, analysis in terms of each of the alternative scales is warranted provided that each of the scales satisfies the assumptions underlying the respective analyses.

It may happen that within-cell variances will be homogeneous in terms of one scale of measurement but heterogeneous in terms of a second scale. The within-cell distributions may be highly skewed in terms of one scale but approximately normal in terms of a second scale. In terms of one scale of measurement, an additive (i.e., no interaction terms) model may be appropriate, whereas in terms of a second scale the additive model will not be appropriate.

In determining the choice of a scale of measurement for the observed data, two cases will be contrasted. In one case, a priori theory and experience determine the appropriate model as well as the appropriate scale. In the second case, where there is neither adequate theory nor experience to serve as guides, the appropriate model and the proper scale of measurement are determined only after the experimental data have been partially

analyzed. In the latter case the design of the experiment should provide the experimenter with sufficient data to permit the evaluation of alternative formulations of the model.

A readable summary of methods for determining appropriate transformations will be found in Olds et al. (1956). A series of alternative methods is also given by Tukey (1949). A transformation in this context is a change in the scale of measurement for the criterion. For example, rather than time in seconds the scale of measurement may be logarithm time in seconds; rather than number of errors, the square root of the number of errors may be used as the criterion score. There are different reasons for making such transformations. Some transformations have as their primary purpose the attainment of homogeneity of error variance. The work of Box (1953) has shown that the distribution of the F ratio in the analysis of variance is affected relatively little by inequalities in the variances which are pooled into the experimental error. Transformations which have homogeneity of error variance as their primary purpose are relatively less important than they were formerly considered to be. With regard to the usual tests for homogeneity of error variance, Box (1953, p. 333) says, "To make the preliminary tests on variances is rather like putting to sea in a rowing boat to find out whether conditions are sufficiently calm for an ocean liner to leave port."

Another reason for using transformations is to obtain normality of within-cell distributions. Often non-normality and heterogeneity of variance occur simultaneously. The same transformation will sometimes normalize the distributions as well as make the variances more homogeneous. The work of Box (1953) has shown that the sampling distribution of the F ratio is relatively insensitive to moderate departures from normality. Hence transformations whose primary purpose is to attain normal within-cell distributions are now considered somewhat less important than was the case previously.

A third reason for transformations is to obtain additivity of effects. In this context *additivity of effects* implies a model which does not contain interaction terms. In some of the designs which are discussed in later chapters, a strictly additive model is required. In designs of this type certain interaction effects are completely confounded with experimental error. For those designs which do permit independent estimation of interaction effects and error effects, the strictly additive model is not essential. There are, however, advantages to the strictly additive model, if it is appropriate, over the nonadditive model—particularly in cases in which fixed and random factors are in the same experiment. The interaction of the random factors with the fixed factors, if these exist, will tend to increase the variance of the sampling distribution of the main effects. Tukey (1949) has pointed out rather vividly the influence of the scale of measurement upon existence or nonexistence of interaction effects.

The use of transformations in order to obtain additivity has received more attention in recent works than it has in the past. Tukey's test for non-additivity has been used, in part, as a guide for deciding between alternative possible transformations. It is not possible to find transformations which will eliminate nonadditivity in all cases. In some cases there is an intrinsic interaction between the factors which cannot be considered a function of the choice of the scale of measurement. These cases are not always easily distinguished from cases in which the interaction is essentially an artifact of the scale of measurement.

A *monotonic* transformation is one which leaves ordinal relationships (i.e., greater than, equal to, or less than) unchanged. If the means for the levels of factor A have the same rank order for all levels of factor B , then a monotonic transformation can potentially remove the $A \times B$ interaction. When such rank order is not present, a monotonic transformation cannot remove the $A \times B$ interaction. Only monotonic transformations will be discussed in this section. Nonmonotonic transformations in this connection would be of extremely limited utility; the author is aware of no studies in which the latter class of transformation has been used.

There are some over-all guides in selecting a scale of measurement which will satisfy the assumptions of homogeneity of error variance. These guides will be considered in terms of the relationship between the cell means and the cell variances. The latter relationship will be presented in terms of a $p \times q$ factorial experiment; the principles to be discussed hold for all designs.

Case (i): $\sigma_{ij}^2 = \mu_{ij}$. In this case the cell variances tend to be functions of the cell means: the larger the mean, the larger the variance. This kind of relationship exists when the within-cell distribution is Poisson in form. For this case, a square-root transformation will tend to make the variances more homogeneous. This transformation has the form

$$X'_{ijk} = \sqrt{X_{ijk}},$$

where X is the original scale and X' is the transformed scale. If X is a frequency, i.e., number of errors, number of positive responses, and if X is numerically small in some cases (say, less than 10), then a more appropriate transformation is

$$X'_{ijk} = \sqrt{X_{ijk}} + \sqrt{X_{ijk} + 1}.$$

The following transformation is also used for frequency data in which some of the entries are numerically small:

$$X'_{ijk} = \sqrt{X_{ijk} + \frac{1}{2}}.$$

Either of the last two transformations is suitable for the stated purpose.

Case (ii): $\sigma_{ij}^2 = \mu_{ij}(1 - \mu_{ij})$. This case occurs in practice when the basic observations have a binomial distribution. For example, if the basic observations are proportions, variances and means will be related in the manner indicated. The following transformation is effective in stabilizing the variances,

$$X'_{ijk} = 2 \arcsin \sqrt{X_{ijk}},$$

where X_{ijk} is a proportion. In many cases only a single proportion appears in a cell. Tables are available for this transformation. Numerically, X'_{ijk} is an angle measured in radians. For proportions between .001 and .999, X'_{ijk} assumes values between .0633 and 3.0783. The notation \sin^{-1} (read inverse sine) is equivalent to the notation \arcsin . For values of X close to zero or close to unity, the following transformation is recommended,

$$X'_{ijk} = 2 \arcsin \sqrt{X_{ijk} + (\frac{1}{2}n)},$$

where n is the number of observations on which X is based.

Case (iii): $\sigma_{ij}^2 = k^2 \mu_{ij}$. In this case the logarithmic transformation will stabilize the variances.

$$X'_{ijk} = \log X_{ijk}.$$

To avoid values of X close to zero, an alternative transformation

$$X'_{ijk} = \log (X_{ijk} + 1)$$

is often used when some of the measurements are equal to or close to zero. The logarithmic transformation is particularly effective in normalizing distributions which have positive skewness. Such distributions occur in psychological research when the criterion is in terms of a time scale, i.e., number of seconds required to complete a task.

Use of the Range in Deciding between Alternative Transformations. In deciding which one of several possible transformations to use in a specific problem, one may investigate several before deciding which one puts the data in a form that most nearly satisfies the basic assumptions underlying the analysis of variance. The use of the range statistic (or the truncated range statistic) provides a relatively simple method for inspecting the potential usefulness of several transformations with a minimum of computational effort. An example given by Rider et al. (1956, pp. 47-55) will be used to illustrate the method.

In this example, eight operators individually measured the resistance of each of four propeller blades with each of two instruments. Order was randomized. The range of the 16 measurements on each blade in terms of the original scale as well as in terms of transformed scales is given below. (Only the end points of the ranges in terms of the original scale need to be transformed in order to obtain the following data.)

Blade	Scale of measurement			
	Original	Square root	Logarithm	Reciprocal
1	3.10	0.61	0.21	0.077
2	0.10	0.08	0.12	0.833
3	0.15	0.12	0.17	1.111
4	11.00	1.01	0.16	0.015

The logarithmic transformation is seen to make the ranges more uniform. In many practical cases the range tends to be proportional to the variance. A transformation which tends to make the ranges uniform will also tend to make the variances more uniform. As a further step in checking on the adequacy of the logarithmic transformation, the authors applied Tukey's test for nonadditivity to each of the four interaction terms. On the original scale of measurement, two of the four F ratios for nonadditivity were significant at the 5 per cent level. None of the F ratios for nonadditivity was significant at the 5 per cent level in terms of the logarithmic scale.

5.22 Unequal Cell Frequencies

Although an experimental design in its initial planning phases may call for an equal number of observations per cell, the completed experiment may not have an equal number of observations in all cells. There may be many reasons for such a state of affairs. The experimenter may be forced to work with intact groups having unequal size; the required number of individuals in a given category may not be available to the experimenter at a specified time; subjects may not show up to complete their part in an experiment; laboratory animals may die in the course of an experiment. If the original plan for an experiment calls for an equal number of observations in each cell, and if the loss of observations in cells is essentially random (in no way directly related to the experimental variables), then the experimental data may appropriately be analyzed by the method of unweighted means. In essence the latter method considers each cell in the experiment as if it contained the same number of observations as all other cells (at least with regard to the computation of main effects and interaction effects).

Under the conditions that have been specified, the number of observations within each cell will be of the same order of magnitude. The procedures for an unweighted-means analysis will be described in terms of a 2×3 factorial experiment. These procedures may be generalized to higher-order factorial experiments. The number of observations in each cell may be indicated as follows:

	b_1	b_2	b_3
a_1	n_{11}	n_{12}	n_{13}
a_2	n_{21}	n_{22}	n_{23}

The harmonic mean of the number of observations per cell is

$$\bar{n}_h = \frac{pq}{(1/n_{11}) + (1/n_{12}) + \cdots + (1/n_{pq})}.$$

In the computation of main effects and interactions each cell is considered to have \bar{n}_h observations. (The harmonic mean rather than the arithmetic mean is used here because the standard error of a mean is proportional to $1/n_{ij}$ rather than n_{ij} .)

The mean for each of the cells may be represented as follows:

	b_1	b_2	b_3
a_1	\overline{AB}_{11}	\overline{AB}_{12}	\overline{AB}_{13}
a_2	\overline{AB}_{21}	\overline{AB}_{22}	\overline{AB}_{23}

The estimate of the mean $\mu_{.1}$ is

$$\bar{A}_1 = \frac{\sum_j \overline{AB}_{1j}}{q}.$$

That is, \bar{A}_1 is the mean of the means in row a_1 , not the mean of all observations at level a_1 . These two means will differ when each cell does not have the same number of observations. The estimate of $\mu_{.1}$ is

$$\bar{B}_1 = \frac{\sum_i \overline{AB}_{i1}}{p}.$$

Again there will be a difference between the mean of the means within a column and the mean of all observations in a column. The grand mean in the population is estimated by

$$G = \frac{\sum_i \bar{A}_i}{p} = \frac{\sum_j \bar{B}_j}{q} = \frac{\sum \sum \overline{AB}_{ij}}{pq}.$$

Variation due to main effects and interactions are estimated by the following sums of squares:

$$SS_a = \bar{n}_h q \sum (\bar{A}_i - \bar{G})^2,$$

$$SS_b = \bar{n}_h p \sum (\bar{B}_j - \bar{G})^2,$$

$$SS_{ab} = \bar{n}_h \sum (\overline{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G})^2.$$

These sums of squares have the same form as corresponding sums of squares for the case of equal cell frequencies. However, \bar{A}_i , \bar{B}_j , and \bar{G} are computed in a different manner. If all cell frequencies were equal, both computational procedures would lead to identical results.

The variation with cell ij is

$$SS_{ij} = \sum_m X_{ijm}^2 - \frac{\left(\sum_m X_{ijm} \right)^2}{n_{ij}}.$$

The pooled within-cell variation is

$$SS_{w, cell} = \sum \sum SS_{ij}.$$

The degrees of freedom for this latter source of variation are

$$df_{w, cell} = (\sum \sum n_{ij}) - pq.$$

Other methods are available for handling the analysis for unequal cell frequencies. If, however, the differences in cell frequencies are primarily functions of sources of variation irrelevant to the experimental variables, there are no grounds for permitting such frequencies to influence the estimation of population means. On the other hand, should the cell frequencies be directly related to the size of corresponding population strata, then such frequencies should be used in estimating the mean of the population composed of such strata.

5.23 Unequal Cell Frequencies—Least-squares Solution

The general least-squares solution for the sums of squares in a factorial experiment is computationally more difficult than the corresponding solution for the unweighted-means analysis. There is some evidence to indicate that the resulting tests, in the least-squares case, are the more powerful. Suppose that the following linear model is appropriate for the first stage of the estimation process:

$$(1) \quad \overline{AB}_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}.$$

The parameters on the right-hand side of (1) satisfy the structural requirements of linear model I. For the case of unequal cell frequencies, the restrictions on the parameters may be taken to be

$$\sum n_{i.} \alpha_i = 0 \quad \text{and} \quad \sum n_{.j} \beta_j = 0.$$

Let the least-squares estimators of the corresponding parameters be designated by $\hat{\mu}$, $\hat{\alpha}_i$, and $\hat{\beta}_j$. These estimators are determined in such a manner as to make

$$(2) \quad \sum \sum (\overline{AB}_{ij} - \hat{\mu} - \hat{\alpha}_i - \hat{\beta}_j)^2 = \text{minimum},$$

subject to the restrictions

$$\sum n_{i.} \hat{\alpha}_i = 0 \quad \text{and} \quad \sum n_{.j} \hat{\beta}_j = 0.$$

The restrictions in (1) and (2) are introduced for computational convenience; they have no effect upon the resulting sums of squares. The same sums of squares would result if the restrictions had the form

$$\sum \hat{\alpha}_i = 0 \quad \text{and} \quad \sum \hat{\beta}_j = 0.$$

By differentiating (2) with respect to $\hat{\mu}$, the least-squares equation defining $\hat{\mu}$ is

$$(3) \quad \hat{\mu} = \frac{G}{n_{..}}.$$

The least-squares equations defining the $\hat{\beta}_j$ can be shown (Kempthorne, 1952, p. 80) to have the following form:

$$(4) \quad \begin{aligned} n'_{11}\hat{\beta}_1 + n'_{12}\hat{\beta}_2 + \cdots + n'_{1q}\hat{\beta}_q &= B'_1, \\ n'_{21}\hat{\beta}_1 + n'_{22}\hat{\beta}_2 + \cdots + n'_{2q}\hat{\beta}_q &= B'_2, \\ &\vdots \\ n'_{q1}\hat{\beta}_1 + n'_{q2}\hat{\beta}_2 + \cdots + n'_{qq}\hat{\beta}_q &= B'_q. \end{aligned}$$

The set of Eqs. (4) is analogous to the set of normal equations encountered in multiple regression. In the present case, the $\hat{\beta}_j$'s have the role of regression coefficients, the n'_{ij} 's the role of covariances among the predictors, and the B'_j 's the role of covariances between predictor and criterion.

The coefficients of the $\hat{\beta}_j$'s are defined as follows:

$$\begin{aligned} n'_{kk} &= n_{kk} - \left[\frac{n_{1k}^2}{n_{1.}} + \frac{n_{2k}^2}{n_{2.}} + \cdots + \frac{n_{pk}^2}{n_{p.}} \right], \\ n'_{km} &= - \left[\frac{n_{1k}n_{1m}}{n_{1.}} + \frac{n_{2k}n_{2m}}{n_{2.}} + \cdots + \frac{n_{pk}n_{pm}}{n_{p.}} \right]. \end{aligned}$$

Since this last expression is symmetric in k and m ,

$$n'_{km} = n'_{mk}.$$

The terms on the right-hand side of (4) are defined by

$$B'_j = B_j - \left[\frac{n_{1j}A_1}{n_{1.}} + \frac{n_{2j}A_2}{n_{2.}} + \cdots + \frac{n_{pj}A_p}{n_{p.}} \right].$$

B'_j is called the adjusted total for the observations at level b_j .

Several procedures exist for solving the set of Eqs. (4) for the $\hat{\beta}_j$'s. It is not, however, necessary to obtain the individual $\hat{\beta}_j$'s in order to obtain the sum of squares associated with the $\hat{\beta}_j$'s. An example of a computational procedure for obtaining this sum of squares in this manner is given in Sec. 6.14. If the individual $\hat{\beta}_j$'s are obtained, then

$$(5) \quad SS_{b(\text{adj})} = \sum \hat{\beta}_j B'_j.$$

A general method for obtaining the $\hat{\beta}_j$'s is described by Kempthorne (1952, p. 81).

A set of simultaneous equations defining the $\hat{\alpha}_i$'s has a form which is similar to (4), i.e.,

$$n'_{i1}\hat{\alpha}_1 + n'_{i2}\hat{\alpha}_2 + \cdots + n'_{ip}\hat{\alpha}_i = A'_i, \quad i = 1, \dots, p.$$

In this case,

$$\begin{aligned} n'_{kk} &= n_{kk} - \left[\frac{n_{k1}^2}{n_{.1}} + \frac{n_{k2}^2}{n_{.2}} + \cdots + \frac{n_{kq}^2}{n_{.q}} \right], \\ n'_{km} &= - \left[\frac{n_{k1}n_{m1}}{n_{.1}} + \frac{n_{k2}n_{m2}}{n_{.2}} + \cdots + \frac{n_{kq}n_{mq}}{n_{.q}} \right], \\ A'_i &= A_i - \left[\frac{n_{i1}B_1}{n_{.1}} + \frac{n_{i2}B_2}{n_{.2}} + \cdots + \frac{n_{iq}B_q}{n_{.q}} \right]. \end{aligned}$$

However, if $SS_{b(\text{adj})}$ is known, $SS_{a(\text{adj})}$ may be obtained without the use of (5). This can be done as follows.

Define the *unadjusted* sums of squares as

$$\begin{aligned}SS_a &= \Sigma \left(\frac{A_i^2}{n_{i.}} \right) - \frac{G^2}{n_{..}}, \\SS_b &= \Sigma \left(\frac{B_j^2}{n_{.j}} \right) - \frac{G^2}{n_{..}}, \\SS_{\text{cells}} &= \Sigma \Sigma \left(\frac{(AB_{ij})^2}{n_{ij}} \right) - \frac{G^2}{n_{..}}.\end{aligned}$$

The following relationship may be shown to hold:

$$(6) \quad SS_a + SS_{b(\text{adj})} = SS_{a(\text{adj})} + SS_b,$$

Hence, if $SS_{b(\text{adj})}$ has been computed, then

$$SS_{a(\text{adj})} = SS_a + SS_{b(\text{adj})} - SS_b.$$

The minimum value that (2) obtains defines the residual variation among cells, i.e., that part of the variation among the cell means that cannot be predicted from the least-squares estimators of the main effects due to factors A and B . This residual variation is actually that due to the AB interaction. It can be shown that

$$(7) \quad SS_{\text{res from cells}} = SS_{ab(\text{adj})} = SS_{\text{cells}} - SS_a - SS_{b(\text{adj})}.$$

Equivalently,

$$SS_{\text{res from cells}} = SS_{ab(\text{adj})} = SS_{\text{cells}} - SS_{a(\text{adj})} - SS_b.$$

If the $\hat{\alpha}_i$'s and $\hat{\beta}_j$'s have been computed, then the interaction sum of squares is given by

$$SS_{ab(\text{adj})} = SS_{\text{cells}} - \Sigma \hat{\alpha}_i A_i - \Sigma \hat{\beta}_j B_j.$$

Since $\hat{\alpha}_i$ and $\hat{\beta}_j$ are determined by a procedure which minimizes this quantity, $SS_{ab(\text{adj})}$ is the least-squares estimate of the variation due to interaction.

There are computational methods for obtaining $SS_{ab(\text{adj})}$ without having either the $\hat{\alpha}_i$'s or the $\hat{\beta}_j$'s (see Rao, 1952, pp. 96-98). Given $SS_{ab(\text{adj})}$, one has

$$SS_{a(\text{adj})} = SS_{\text{cells}} - SS_{ab(\text{adj})} - SS_b,$$

$$SS_{b(\text{adj})} = SS_{\text{cells}} - SS_{ab(\text{adj})} - SS_a.$$

A numerical example of these computational procedures is given in Sec. 6.14.

The within-cell variation is defined in the same manner as that described in the unweighted-means analysis. Over-all tests follow the same procedure as that used in the case of equal cell frequencies. Should the interaction be significant, primary interest is in the simple main effects rather than the over-all main effects. Analysis of the simple main effects takes the form of a single-classification analysis of variance having unequal cell frequencies.

The complete set of normal equations obtained from (2), for the special case of a 2×3 factorial experiment, may be represented schematically as follows.

$$\begin{bmatrix} \hat{\mu} & \hat{\alpha}_1 & \hat{\alpha}_2 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \\ \Sigma \Sigma n_{ij} & n_{1.} & n_{2.} & n_{.1} & n_{.2} & n_{.3} \\ n_{1.} & n_{1.} & 0 & n_{11} & n_{12} & n_{13} \\ n_{2.} & 0 & n_{2.} & n_{21} & n_{22} & n_{23} \\ n_{.1} & n_{11} & n_{21} & n_{.1} & 0 & 0 \\ n_{.2} & n_{12} & n_{22} & 0 & n_{.2} & 0 \\ n_{.3} & n_{13} & n_{23} & 0 & 0 & n_{.3} \end{bmatrix} = \begin{bmatrix} G \\ A_1 \\ A_2 \\ B_1 \\ B_2 \\ B_3 \end{bmatrix}.$$

The normal equations given by (3) and (4) represent subsets of the above complete set. The subset (3) is obtained by "sweeping out" the effects of factor A . In the above schematic representation, the second equation has the form

$$n_{1.}\hat{\mu} + n_{1.}\hat{\alpha}_1 + 0\hat{\alpha}_2 + n_{11}\hat{\beta}_1 + n_{12}\hat{\beta}_2 + n_{13}\hat{\beta}_3 = A_1.$$

CHAPTER 6

Factorial Experiments—Computational Procedures and Numerical Examples

6.1 General Purpose

In this chapter the principles discussed in Chap. 5 will be illustrated by numerical examples; detailed computational procedures will be given for a variety of factorial experiments. It should be noted that the formulas convenient for computational work are not necessarily those which are most directly interpretable. For purposes of interpretation the basic definitions given in Chap. 5 are the important sources for reference; however, in all cases the computational formulas are algebraically equivalent to the basic definitions. The algebraic proofs underlying this equivalence are not difficult. Factorial experiments in which there are repeated measures are discussed in Chap. 7.

6.2 $p \times q$ Factorial Experiment Having n Observations per Cell

The treatment combinations in this type of experiment are represented in the cells of the following table:

	b_1	b_2	\cdots	b_j	\cdots	b_q
a_1	ab_{11}	ab_{12}	\cdots	ab_{1j}	\cdots	ab_{1q}
a_2	ab_{21}	ab_{22}	\cdots	ab_{2j}	\cdots	ab_{2q}
\vdots						
a_i	ab_{i1}	ab_{i2}	\cdots	ab_{ij}	\cdots	ab_{iq}
\vdots						
a_p	ab_{p1}	ab_{p2}	\cdots	ab_{pj}	\cdots	ab_{pq}

The cell in row a_i and column b_j corresponds to that part of the experiment in which treatment a_i is used in combination with b_j to yield the treatment

combination ab_{ij} . The first subscript in this notation system refers to the level of factor A , the second subscript to the level of factor B .

Within each cell of the experiment there are n observations. The observations made under treatment combination ab_{ij} may be symbolized as follows:

$$a_i \quad \boxed{X_{ij1} X_{ij2} \cdots X_{ijk} \cdots X_{ijn}} \quad b_j$$

The symbol X_{ij1} denotes a measurement on the first element in this cell. The symbol X_{ij2} denotes a measurement on the second element in this cell. The measurement on element k is denoted by the symbol X_{ijk} . The subscript k assumes the values $1, 2, \dots, n$ within each of the cells. The sum of the n observations within cell ab_{ij} will be denoted by the symbol AB_{ij} ; thus,

$$AB_{ij} = \sum_k X_{ijk}.$$

The following table summarizes the notation that will be used for sums of basic measurements:

	b_1	b_2	\cdots	b_j	\cdots	b_q	Row sum = \sum_j
a_1	AB_{11}	AB_{12}	\cdots	AB_{1j}	\cdots	AB_{1q}	$A_1 = \sum_j AB_{1j}$
a_2	AB_{21}	AB_{22}	\cdots	AB_{2j}	\cdots	AB_{2q}	A_2
\vdots							
a_i	AB_{i1}	AB_{i2}	\cdots	AB_{ij}	\cdots	AB_{iq}	A_i
\vdots							
a_p	AB_{p1}	AB_{p2}	\cdots	AB_{pj}	\cdots	AB_{pq}	A_p
Column sum	B_1	B_2	\cdots	B_j	\cdots	B_q	G

The sum of the nq measurements in row i , that is, the sum of all measurements made at level a_i , is denoted by the symbol A_i . Thus,

$$A_i = \sum_j AB_{ij} = \sum_j \sum_k X_{ijk}.$$

The double summation symbol $\sum_j \sum_k$ indicates that one sums within each cell as well as across all cells in row i . (Summing over the subscripts j and k is equivalent to summing over all observations within a given row.) The sum of the np measurements in column j , that is, the sum of all measurements made under level b_j , is denoted by the symbol B_j . Thus,

$$B_j = \sum_i AB_{ij} = \sum_i \sum_k X_{ijk}.$$

The grand total of all measurements is denoted by the symbol G . Thus,

$$G = \sum_i A_i = \sum_j B_j = \sum_i \sum_j AB_{ij} = \sum_i \sum_j \sum_k X_{ijk}.$$

The mean of all measurements under treatment combination ab_{ij} is

$$\overline{AB}_{ij} = \frac{AB_{ij}}{n}.$$

The mean of all measurements at level a_i is

$$\bar{A}_i = \frac{A_i}{nq}.$$

The mean of all measurements at level b_j is

$$\bar{B}_j = \frac{B_j}{np}.$$

The grand mean of all observations in the experiment is

$$\bar{G} = \frac{G}{npq}.$$

To summarize the notation, a capital letter with a subscript ijk represents an individual observation; a pair of capital letters with the subscript ij represents the sum over the n observations represented by the subscript k . A capital letter with the subscript i represents the sum of nq observations within a row of the experimental plan; a capital letter with the subscript j represents the sum of the np observations within a column of the experimental plan. A widely used equivalent notation system is summarized below:

<i>Notation</i>	<i>Equivalent notation</i>
X_{ijk}	X_{ijk}
AB_{ij}	$X_{ij..}$
A_i	$X_{i...}$
B_j	$X_{.j.}$
G	$X_{...}$

In the equivalent notation system, the periods indicate the subscript over which the summation has been made.

Definition of Computational Symbols. In order to simplify the writing of the computational formulas for the sums of squares needed in the analysis of variance, it is convenient to introduce a set of computational symbols. The symbols appropriate for a $p \times q$ factorial experiment having n observations per cell are defined in Table 6.2-1. The computational procedures for this case are more elaborate than they need be, but the procedures to be developed here can be readily extended to more complex experimental plans in which they are not more elaborate than they need be.

In using the summation notation, where the index of summation is not

indicated it will be understood that the summation is over all possible subscripts. For example, the notation $\sum_i \sum_j \sum_k X_{ijk}^2$ will be abbreviated ΣX_{ijk}^2 ; similarly the notation $\sum_i \sum_j (AB_{ij})^2$ will be abbreviated $\Sigma (AB_{ij})^2$. Where the summation is restricted to a single level of one of the subscripts, the index of summation will be indicated. Thus,

$$\sum_i (AB_{ij})^2 = AB_{1j}^2 + AB_{2j}^2 + \cdots + AB_{ij}^2 + \cdots + AB_{pj}^2,$$

$$\sum_j (AB_{ij})^2 = AB_{i1}^2 + AB_{i2}^2 + \cdots + AB_{ij}^2 + \cdots + AB_{iq}^2.$$

In the definitions of the computational formulas given in part i of Table 6.2-1, the divisor in each case is the number of observations summed to

Table 6.2-1 Definition of Computational Symbols

(i)	(1) = G^2/npq	(3) = $(\Sigma A_i^2)/nq$
	(2) = ΣX_{ijk}^2	(4) = $(\Sigma B_j^2)/np$
	(5) = $[\Sigma (AB_{ij})^2]/n$	
(ii)	Source of variation	Computational formula for SS
	A	$SS_a = (3) - (1)$
	B	$SS_b = (4) - (1)$
	AB	$SS_{ab} = (5) - (3) - (4) + (1)$
	Experimental error (within cell)	$SS_{\text{error}} = (2) - (5)$
	Total	$SS_{\text{total}} = (2) - (1)$

obtain one of the terms that is squared. For example, n observations are summed to obtain an AB_{ij} . Hence the denominator n in the term $[\Sigma (AB_{ij})^2]/n$. There are nq observations summed to obtain an A_i . Hence the denominator nq in the term $(\Sigma A_i^2)/nq$.

An estimate of the variation due to the main effects of factor A is given by

$$SS_a = nq \Sigma (\bar{A}_i - \bar{G})^2.$$

This is not a convenient computational formula for this source of variation. An algebraically equivalent form is

$$SS_a = \frac{\Sigma A_i^2}{nq} - \frac{G^2}{npq} = (3) - (1).$$

The other sources of variation in the analysis of variance are as follows:

$$SS_b = np \Sigma (\bar{B}_j - \bar{G})^2 = (4) - (1),$$

$$SS_{ab} = n(\Sigma \bar{A}_i \bar{B}_j - \bar{A}_i - \bar{B}_j + \bar{G})^2 = (5) - (3) - (4) + (1),$$

$$SS_{\text{w. cell}} = \Sigma (X_{ijk} - \bar{AB}_{ij})^2 = (2) - (5),$$

$$SS_{\text{total}} = \Sigma (X_{ijk} - \bar{G})^2 = (2) - (1).$$

These computational formulas are summarized in part ii of Table 6.2-1.

Should the interaction term in the analysis of variance prove to be statistically significant, it is generally necessary to analyze the simple main effects rather than the over-all main effects. Computational symbols for variation due to simple main effects are summarized in Table 6.2-2. By definition the

Table 6.2-2 Definition of Computational Symbols for Simple Effects

(i)	$(3a_1) = A_1^2/nq$	$(4b_1) = B_1^2/np$
	$(3a_2) = A_2^2/nq$	$(4b_2) = B_2^2/np$

	$(3a_p) = A_p^2/nq$	$(4b_q) = B_q^2/np$
	$(3) = (\Sigma A_i^2)/nq$	$(4) = (\Sigma B_j^2)/np$
	$(5a_1) = [\Sigma_j (AB_{1j})^2]/n$	$(5b_1) = [\Sigma_i (AB_{i1})^2]/n$
	$(5a_2) = [\Sigma_j (AB_{2j})^2]/n$	$(5b_2) = [\Sigma_i (AB_{i2})^2]/n$

	$(5a_p) = [\Sigma_j (AB_{pj})^2]/n$	$(5b_q) = [\Sigma_i (AB_{iq})^2]/n$
	$(5) = [\Sigma (AB_{ij})^2]/n$	$(5) = [\Sigma (AB_{ij})^2]/n$
		Computational formula for SS
(ii)	Simple effects for A :	
	For level b_1	$(5b_1) - (4b_1)$
	For level b_2	$(5b_2) - (4b_2)$

	For level b_q	$(5b_q) - (4b_q)$
	Simple effects for B :	
	For level a_1	$(5a_1) - (3a_1)$
	For level a_2	$(5a_2) - (3a_2)$

	For level a_p	$(5a_p) - (3a_p)$

variation due to the simple main effect of factor A for level b_1 is

$$\begin{aligned} SS_{a \text{ for } b_1} &= n \Sigma (\overline{AB_{1j}} - \bar{B}_1)^2 \\ &= \frac{\Sigma (AB_{1j})^2}{n} - \frac{B_1^2}{np} = (5b_1) - (4b_1). \end{aligned}$$

By definition, the variation due to the simple main effect of factor B for level a_1 is

$$\begin{aligned} SS_{b \text{ for } a_1} &= n \Sigma (\overline{AB_{1j}} - \bar{A}_1)^2 \\ &= \frac{\Sigma (AB_{1j})^2}{n} - \frac{A_1^2}{nq} = (5a_1) - (3a_1). \end{aligned}$$

Numerical Example. A numerical example of a 2×3 factorial experiment having three observations per cell will be used to illustrate the computational procedures. Suppose that an experimenter is interested in evaluating the relative effectiveness of three drugs (factor B) in bringing about behavioral changes in two categories, schizophrenics and depressives,

Table 6.2-3 Numerical Example

Observed data:

	Drug b_1			Drug b_2			Drug b_3		
(i) Category a_1	8	4	0	10	8	6	8	6	4
Category a_2	14	10	6	4	2	0	15	12	9

AB summary table:

	b_1	b_2	b_3	Total
(ii) a_1	12	24	18	$54 = A_1$
a_2	30	6	36	$72 = A_2$
	42	30	54	$126 = G$
	B_1	B_2	B_3	

(iii)	(1) = $(126)^2/18$	= 882.00
	(2) = $(8^2 + 4^2 + 0^2 + \dots + 15^2 + 12^2 + 9^2)/9$	= 1198
	(3) = $(54^2 + 72^2)/9$	= 900.00
	(4) = $(42^2 + 30^2 + 54^2)/6$	= 930.00
	(5) = $(12^2 + 24^2 + 18^2 + 30^2 + 6^2 + 36^2)/3$	= 1092.00
<hr/>		
	$SS_a = (3) - (1) = 900.00 - 882.00$	= 18.00
	$SS_b = (4) - (1) = 930.00 - 882.00$	= 48.00
(iv)	$SS_{ab} = (5) - (3) - (4) + (1)$	
	$= 1092.00 - 900.00 - 930.00 + 882.00$	= 144.00
	$SS_{w, cell} = (2) - (5) = 1198 - 1092.00$	= 106.00
	$SS_{total} = (2) - (1) = 1198 - 882.00$	= 316.00

of patients (factor A). What is considered to be a random sample of nine patients belonging to category a_1 (schizophrenics) is divided at random into three subgroups, with three patients in each subgroup. Each subgroup is then assigned to one of the drug conditions. An analogous procedure is followed for a random sample of nine patients belonging to category a_2 (depressives). Criterion ratings are made on each patient before and after the administration of the drugs. The numerical entries in part i of Table 6.2-3 represent the difference between the two ratings on each of the patients. (An analysis of covariance might be more appropriate for this plan; covariance analysis is discussed in Chap. 11.)

As a first step in the analysis, the AB summary table in part ii is obtained. The entry in row a_1 , column b_1 is

$$AB_{11} = 8 + 4 + 0 = 12.$$

The entry in row a_2 , column b_3 is

$$AB_{23} = 15 + 12 + 9 = 36.$$

Data for all computational symbols except (2) are in part ii; computational symbol (2) is obtained from data in part i. The analysis of variance is summarized in Table 6.2-4.

Table 6.2-4 Summary of Analysis of Variance

Source of variation	SS	df	MS	F
A (category of patient)	18.00	1	18.00	2.04
B (drug)	48.00	2	24.00	2.72
AB	144.00	2	72.00	8.15
Within cell	106.00	12	8.83	
Total	316.00	17		

The structure of the F ratios used in making tests depends upon the expected values of mean squares appropriate for the experimental data. If the categories and drugs are fixed factors, i.e., if inferences are to be made only with respect to the two categories of patients represented in the experiment and only with respect to the three drugs included in the experiment, then the appropriate expected values of the mean squares are given in Table 6.2-5. The structure of F ratios is determined in accordance with the principles given in Sec. 5.7.

Table 6.2-5 Expected Values for Mean Squares

Source	MS	$E(MS)$	F
Main effect of A	MS_a	$\sigma_\epsilon^2 + 9\sigma_\alpha^2$	$MS_a/MS_{w, cell}$
Main effect of B	MS_b	$\sigma_\epsilon^2 + 6\sigma_\beta^2$	$MS_b/MS_{w, cell}$
AB interaction	MS_{ab}	$\sigma_\epsilon^2 + 3\sigma_{\alpha\beta}^2$	$MS_{ab}/MS_{w, cell}$
Within cell	$MS_{w, cell}$	σ_ϵ^2	

If tests are made at the .05 level of significance, the critical value for the test of the hypothesis that the action of the drugs is independent of the category of patient (i.e., zero interaction) is $F_{.95}(2, 12) = 3.88$. In this case $F_{obs} = 8.15$ exceeds the critical value. Hence the experimental data do not support the hypothesis of zero interaction. The data indicate that the effect of a drug differs for the two types of patients—the effect of a drug depends upon the category of patient to which it is administered. A

significant interaction indicates that a given drug has different effects for one category of patient from what it has for a second category. The nature of the interaction effects is indicated by inspecting of the cell means. These means are given below:

	Drug 1	Drug 2	Drug 3
Category 1	4	8	6
Category 2	10	2	12

A geometric representation of these means is given in Fig. 6.1. This figure represents the profiles corresponding to the simple effects of the drugs (factor B) for each of the categories (factor A). A test for the presence of

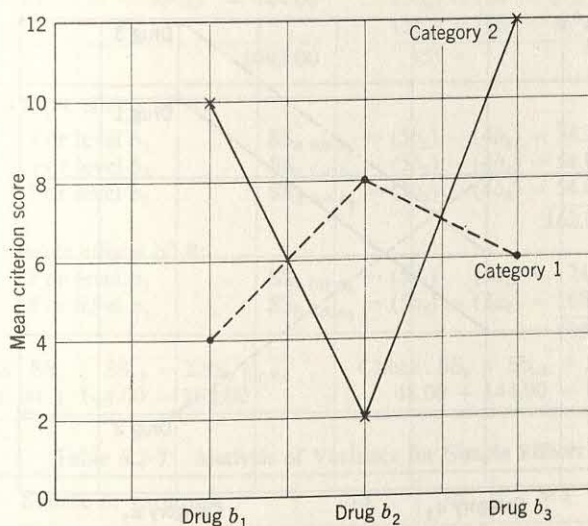


Figure 6.1 Profiles of simple effects for drugs.

interaction is equivalent to a test on the difference in the shapes of the profiles of these simple effects. An equivalent geometric representation of the table of means is given in Fig. 6.2. This figure represents the profiles corresponding to the simple effects of the categories for each of the drugs. The profile for drug 2 appears to have a slope which is different from the slopes of the profiles for the other drugs. As an aid in the interpretation of interactions, geometric representation of the profiles corresponding to the means is generally of considerable value.

Tests on differences between means within the same profile are given by tests on simple effects. Computational procedures for obtaining the variation due to simple effects are summarized in Table 6.2-6. Data from which the symbols in part ii are obtained are given in part i; the latter is the AB summary of Table 6.2-3.

The analysis of variance for the simple effects of the drugs for each of the categories is summarized in Table 6.2-7. The structure of the F ratios for simple effects is dependent upon appropriate expected values for the mean squares. Assuming that factors A and B are fixed factors, expected values for the mean squares of simple effects are given in Table 6.2-8. It should be noted, in terms of the general linear model, that a simple main effect is actually a sum of an over-all main effect and an interaction. For example,

$$\overline{AB}_{ij} - \bar{B}_j \quad \text{estimates} \quad \alpha_i + \alpha\beta_{ij} = \alpha_i \text{ for } b_j.$$

A test on the simple effect of factor A for level b_j is equivalent to a test that $\sigma_{\alpha}^2 \text{ for } b_j = 0$. The appropriate F ratio for this test would have an estimate of σ_{ϵ}^2 as denominator; the latter estimate is given by $MS_{w, \text{ cell}}$.

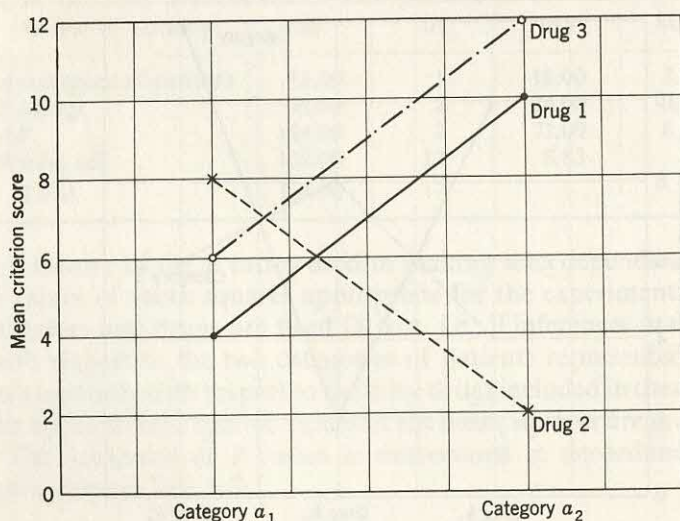


Figure 6.2 Profiles of simple effects for categories.

By using the .05 level of significance for the data in Table 6.2-7, the critical value for an F ratio is $F_{.95}(2,12) = 3.88$. The experimental data indicate no difference between the drugs for category a_1 ($F_{\text{obs}} = 1.36$). The experimental data indicate that there are differences between the drugs for category a_2 ($F_{\text{obs}} = 9.51$). Inspection of the profile of the drugs for category a_2 in Fig. 6.2 indicates that the effect of drug b_2 is different from the effects of the other drugs.

Returning to the analysis in Table 6.2-4, neither of the main effects is statistically significant. However, in the presence of interaction, inferences made with respect to main effects must be interpreted with caution. The data indicate that the main effects due to drugs do not differ. However, this does not mean that the drugs are equally effective for each of the categories considered separately. Because of the significant interaction effect,

Table 6.2-6 Computation of Simple Effects

AB summary table:

12 Summary Table

	b_1	b_2	b_3	Total
(i) a_1	12	24	18	54
a_2	30	6	36	72
	42	30	54	126

$(3a_1) = (54^2)/9 = 324.00$
 $(3a_2) = (72^2)/9 = 576.00$

$(4b_1) = (42^2)/6 = 294.00$
 $(4b_2) = (30^2)/6 = 150.00$
 $(4b_3) = (54^2)/6 = 486.00$

(ii) (3) =	900.00	(4) =	930.00
$(5a_1) = (12^2 + 24^2 + 18^2)/3 = 348.00$		$(5b_1) = (12^2 + 30^2)/3 = 348.00$	
$(5a_2) = (30^2 + 6^2 + 36^2)/3 = 744.00$		$(5b_2) = (24^2 + 6^2)/3 = 204.00$	
		$(5b_3) = (18^2 + 36^2)/3 = 540.00$	
(5) =	1092.00	(5) =	1092.00

Simple effects of *A*:

For level b_1	$SS_{a \text{ for } b_1} = (5b_1) - (4b_1) = 54.00$
For level b_2	$SS_{a \text{ for } b_2} = (5b_2) - (4b_2) = 54.00$
For level b_3	$SS_{a \text{ for } b_3} = (5b_3) - (4b_3) = 54.00$
	162.00

(iii) Simple effects of *B*:

For level a_1	$SS_{b \text{ for } a_1} = (5a_1) - (3a_1) = 24.00$
For level a_2	$SS_{b \text{ for } a_2} = (5a_2) - (3a_2) = 168.00$
	192.00

Check: $SS_a + SS_{ab} = \Sigma SS_{a \text{ for } b_j}$
 $18.00 + 144.00 = 162.00$

Check: $SS_b + SS_{ab} = \Sigma SS_{b \text{ for } a_i}$
 $48.00 + 144.00 = 192.00$

Table 6.2-7 Analysis of Variance for Simple Effects

Source of variation	SS	df	MS	<i>F</i>
<i>B</i> for a_1 (drugs for category a_1)	24.00	2	12.00	1.36
<i>B</i> for a_2 (drugs for category a_2)	168.00	2	84.00	9.51
Within cell	106.00	12	8.83	

Table 6.2-8 Expected Values for Mean Squares of Simple Effects

Source of variation	$E(MS)$
Simple effects of factor <i>A</i> :	
For level b_1	$\sigma_\epsilon^2 + 3\sigma_\alpha^2 \text{ for } b_1$
For level b_2	$\sigma_\epsilon^2 + 3\sigma_\alpha^2 \text{ for } b_2$
For level b_3	$\sigma_\epsilon^2 + 3\sigma_\alpha^2 \text{ for } b_3$
Simple effects for factor <i>B</i> :	
For level a_1	$\sigma_\epsilon^2 + 3\sigma_\beta^2 \text{ for } a_1$
For level a_2	$\sigma_\epsilon^2 + 3\sigma_\beta^2 \text{ for } a_2$

conclusions with respect to main effects due to drugs cannot be applied separately to each of the categories.

The drugs are differentially effective only when one considers comparisons within each of the categories separately. It must be noted and emphasized that statistical inference is a tool to help the experimenter in drawing scientifically meaningful conclusions from his experiment. Too frequently the tool is allowed to become the master rather than the servant. The primary objective of the experiment should determine which tests are to be made and which tests are to be avoided for lack of meaning. Statistical elegance does not necessarily imply scientifically meaningful inferences.

Individual Comparisons. To make comparisons between two or more means, the procedures given in Sec. 5.17 may be used. For example, to test the hypothesis that drugs 1 and 2 are equally effective for category 2, one may use the following statistic (data from the last numerical example will be used for illustrative purposes):

$$\begin{aligned} F &= \frac{(AB_{21} - AB_{22})^2}{2nMS_{w, cell}} \\ &= \frac{(30 - 6)^2}{6(8.83)} = 10.87. \end{aligned}$$

For a .05-level test, the critical value is $F_{.95}(1,12) = 4.75$. Hence the hypothesis that $\mu_{21} = \mu_{22}$ is not supported by the experimental data.

To test differences between all possible pairs of means in a *logical grouping* of means, the procedures given in Sec. 3.9 may be adapted for use. The Newman-Keuls procedure will be illustrated here. Suppose that it is desired to test the differences between all possible pairs of means for category a_2 . The procedures for such multiple comparisons are outlined in Table 6.2-9.

In part i the means to be compared are first arranged in rank order, from low to high. Then differences between all possible pairs which give a positive value are obtained. An estimate of the standard error of a single mean is computed in part ii. This estimate is based upon data in all cells in the experiment, not just those from which the means are obtained. (If there is any real question about homogeneity of variance, only cells from which the means are obtained may be used in estimating the standard error of a mean.) If the within-cell variance from all cells in the experiment is used, degrees of freedom for s_F are $pq(n-1)$, which in this case is equal to 12.

To obtain critical values for a .05-level test, one obtains values of the $q_{.95}(r,12)$ statistic, where 12 is the degrees of freedom for s_F and r is the number of steps two means are apart in an ordered sequence. These values are obtained from tables of the q statistic in Table B.4. The actual critical values are $s_F q_{.95}(r,12)$. For example, the difference

$$\overline{AB}_{23} - \overline{AB}_{22} = 10.$$

These means are three steps apart in the ordered sequence ($r = 3$); hence the critical value is 6.45. The difference

$$\overline{AB}_{23} - \overline{AB}_{21} = 2,$$

for which $r = 2$, has the critical value 5.27. Tests of this kind must be made in a sequence which is specified in Sec. 3.9.

Table 6.2-9 Comparisons between Means for Category a_2

Ordered means:		\overline{AB}_{22}	\overline{AB}_{21}	\overline{AB}_{23}
		2	10	12
		\overline{AB}_{22}	\overline{AB}_{21}	\overline{AB}_{23}
(i)	Differences:	\overline{AB}_{22}	—	8
		\overline{AB}_{21}	—	2
<hr/>				
(ii) $s_{\sqrt{F}} = \sqrt{MS_{w, cell}/n} = \sqrt{8.83/3} = 1.71$				
<hr/>				
$q_{.95}(r, 12):$		3.08	3.77	
$s_{\sqrt{F}}q_{.95}(r, 12):$		5.27	6.45	
<hr/>				
(iii)		\overline{AB}_{22}	\overline{AB}_{21}	\overline{AB}_{23}
	\overline{AB}_{22}		*	*
	\overline{AB}_{21}			—

From the outcome of the tests in part iii one concludes that, for category a_2 , drugs 1 and 3 differ from drug 2 but that there is no statistically significant difference (on the criterion used) between drugs 1 and 3.

Test for Homogeneity of Error Variance. Although F tests in the analysis of variance are robust with respect to the assumption of homogeneity of error variance, a rough but simple check may be made on this assumption through use of the F_{\max} statistic. Apart from the use of this statistic, the variances of the individual cells should be inspected for any kind of systematic pattern between treatments and variances. In cases having a relatively large number of observations within each cell, within-cell distributions should also be inspected.

Use of the F_{\max} test for homogeneity of variance will be illustrated for the data in part i of Table 6.2-3. For this purpose the within-cell variation for each of the cells is required. In the computational procedures given in Table 6.2-3, the pooled within-cell variation from all cells is computed. The variation within cell ij has the form

$$\begin{aligned} SS_{ij} &= \sum_k (X_{ijk} - \overline{AB}_{ij})^2 \\ &= \sum X_{ijk}^2 - \frac{(\overline{AB}_{ij})^2}{n}. \end{aligned}$$

For example,

$$\begin{aligned} SS_{11} &= (8^2 + 4^2 + 0^2) - \frac{(12)^2}{3} \\ &= 80 - 48 = 32. \end{aligned}$$

Similarly, the variation within cell ab_{21} is given by

$$\begin{aligned} SS_{21} &= (14^2 + 10^2 + 6^2) - \frac{(30)^2}{3} \\ &= 332 - 300 = 32. \end{aligned}$$

The other within-cell sums of squares are

$$SS_{12} = 8, \quad SS_{22} = 8, \quad SS_{13} = 8, \quad SS_{24} = 18.$$

As a check on the computational work,

$$\Sigma SS_{ij} = SS_{w. cell} = 106.$$

Since the number of observations in each cell is constant, the F_{\max} statistic is given by

$$F_{\max} = \frac{SS(\text{largest})}{SS(\text{smallest})} = \frac{32}{8} = 4.00.$$

The critical value for a .05-level test is $F_{.95}(pq, n - 1)$, which in this case is $F_{.95}(6, 2) = 266$. Since the observed F_{\max} statistic does not exceed the critical value, the hypothesis of homogeneity of variance may be considered tenable. In cases in which the assumptions of homogeneity of variance cannot be considered tenable, a transformation on the scale of measurement may provide data which are amenable to the assumptions underlying the analysis model.

Approximate F Tests When Cell Variances Are Heterogeneous. A procedure suggested by Box (1954, p. 300) may be adapted for use in testing simple effects for factor A , even though variances may be heterogeneous. The F ratio in this case has the form

$$F = \frac{MS_{a \text{ for } b_j}}{MS_{\text{error}(b_j)}},$$

where $MS_{\text{error}(b_j)}$ is the pooled within-cell variance for all cells at level b_j . The approximate degrees of freedom for this F ratio are

$$\begin{aligned} 1 & \quad \text{for numerator,} \\ n - 1 & \quad \text{for denominator.} \end{aligned}$$

In testing simple effects for factor B at level a_i , the F ratio has the form

$$F = \frac{MS_{b \text{ for } a}}{MS_{\text{error}(a_i)}},$$

where $MS_{\text{error}(a_i)}$ is the pooled within-cell variance for all cells at level a_i . The approximate degrees of freedom for this F ratio are $(1, n - 1)$. In the usual test (assuming homogeneity of variance) the degrees of freedom for the latter F ratio are $[(q - 1), pq(n - 1)]$ if all cell variances are pooled and $[(q - 1), q(n - 1)]$ if only variances from cells at level a_i are pooled.

Alternative Notation Systems. The notation system that has been adopted for use in this and following sections is not the most widely used system but rather a slight variation on what is essentially a common theme running through several notation systems. Bennett and Franklin (1954) use a closely related notation. The equivalence between the two systems is expressed in terms of the following relationships:

$$\begin{aligned} C &= \frac{G^2}{npq}, \\ C_i &= \frac{\sum A_i^2}{nq}, \\ C_j &= \frac{\sum B_j^2}{np}, \\ C_{ij} &= \frac{\sum (AB_{ij})^2}{n}, \\ C_{ijk} &= \sum X_{ijk}^2. \end{aligned}$$

In terms of the Bennett and Franklin notation system, the sums of squares have the following symmetric form:

$$\begin{aligned} SS_a &= SS_i = C_i - C, \\ SS_b &= SS_j = C_j - C, \\ SS_{ab} &= SS_{ij} = C_{ij} - C_i - C_j + C, \\ SS_{\text{w. cell}} &= C_{ijk} - C_{ij}. \end{aligned}$$

The notation used by Kempthorne (1952) is perhaps the most widely used. In this system a single observation is designated Y_{ijk} .

$$\begin{aligned} CF &= \frac{G^2}{npq}, \\ \sum Y_{i..}^2 &= \sum A_i^2, \\ \sum Y_{.j.}^2 &= \sum B_j^2, \\ \sum_{i,j} Y_{ij.}^2 &= \sum (AB_{ij})^2. \end{aligned}$$

6.3 $p \times q$ Factorial Experiment—Unequal Cell Frequencies

Computational procedures for an unweighted-means analysis will be described in this section. The conditions under which this kind of analysis is appropriate are given in Sec. 5.22. For illustrative purposes the computational procedures are cast in terms of a 2×4 factorial experiment;

these procedures may, however, be generalized to any $p \times q$ factorial experiment. Computational procedures for the least-squares solution are given in Sec. 6.14.

Suppose that the levels of factor A represent two methods for calibrating dials and levels of factor B represent four levels of background illumination. The criterion measure is an accuracy score for a series of trials. The

Table 6.3-1 Numerical Example

Observed data:

	b_1	b_2	b_3	b_4
a_1	3, 4, 6, 7	5, 6, 6, 7, 7	4, 6, 8, 8	8, 10, 10, 7, 11
a_2	2, 3, 4	3, 5, 6, 3	9, 12, 12, 8	9, 7, 12, 11

Cell data:

		b_1	b_2	b_3	b_4
a_1	n_{ij}	4	5	4	5
	ΣX	20	31	26	46
	ΣX^2	110	195	180	434
	SS_{ij}	10.00	2.80	11.00	10.80
a_2	n_{ij}	3	4	4	4
	ΣX	9	17	41	39
	ΣX^2	29	79	433	395
	SS_{ij}	2.00	6.75	12.75	14.75

$$\bar{n}_h = \frac{8}{.25 + .20 + .25 + .20 + .33 + .25 + .25 + .25} = 4.04$$

$$SS_{w, \text{ cell}} = \Sigma \Sigma SS_{ij} = 10.00 + 2.80 + \cdots + 14.75 = 70.85$$

original experiment called for five observations per cell. However, because of conditions not related to the experimental variables, the completed experiment had three to five observations per cell. The observed criterion scores are given in part i of Table 6.3-1. Summary of within-cell information required in the analysis is given in part ii. The variation within cell ab_{11} is

$$SS_{11} = 110 - \frac{(20)^2}{4} = 10.00.$$

The harmonic mean of the cell frequencies is computed in part iii. The computational formula used is

$$\bar{n}_h = \frac{pq}{\Sigma \Sigma (1/n_{ij})}$$

The pooled within-cell variation is also computed in part iii.

The data in the cells of part i of Table 6.3-2 are means of the respective n_{ij} observations in the cells. All the computational symbols in part ii are based upon these means and row and column totals of these means. In

Table 6.3-2 Numerical Example (Continued)

Cell means:

	b_1	b_2	b_3	b_4	Total
a_1	5.00	6.20	6.50	9.20	26.90
a_2	3.00	4.25	10.25	9.75	27.25
	8.00	10.45	16.75	18.95	54.15

(i)

$$(1) = G^2/pq = (54.15)^2/8 = 366.53$$

$$(2) = \Sigma X^2 \text{ (see part ii, Table 6.3-1)}$$

(ii)

$$(3) = (\Sigma \bar{A}_i^2)/q = (26.90^2 + 27.25^2)/4 = 366.54$$

$$(4) = (\Sigma \bar{B}_j^2)/p = (8.00^2 + 10.45^2 + 16.75^2 + 18.95^2)/2 = 406.43$$

$$(5) = \Sigma (\bar{AB}_{ij})^2 = 5.00^2 + 6.20^2 + \dots + 9.75^2 = 417.52$$

(iii)

$$SS_a = \bar{n}_h[(3) - (1)] = 4.04[366.54 - 366.53] = .04$$

$$SS_b = \bar{n}_h[(4) - (1)] = 4.04[406.44 - 366.53] = 161.24$$

$$SS_{ab} = \bar{n}_h[(5) - (3) - (4) + (1)] = 44.72$$

defining the computational symbols in (ii), each of the cell means is considered as if it were a single observation. Computational formulas for the main effects and interaction are given in part iii.

The analysis of variance is summarized in Table 6.3-3. The degrees of

Table 6.3-3 Summary of Analysis of Variance

Source of variation	SS	df	MS	F
A (method of calibration)	.04	1	.04	
B (background illumination)	161.20	3	53.73	18.99
AB	44.72	3	14.91	5.27
Within cell	70.85	25	2.83	

freedom for the within-cell variation are $\Sigma \Sigma n_{ij} - pq = 33 - 8 = 25$. If factors A and B are fixed, then $MS_{w. cell}$ is the proper denominator for all tests. By using the .05 level of significance, the critical value for the test on the interaction is $F_{.95}(3,25) = 2.99$. Since the observed F ratio, $F = 5.27$, is larger than the critical value, the data tend to contradict the hypothesis of zero interaction. The test on the main effect for factor B has the critical value 2.99. The observed F ratio, $F = 18.99$, is larger than the critical value for a .05-level test. Hence the data contradict the hypothesis that the main effects of factor B are zero. Inspection of profiles (Fig. 6.3) of the simple effects of B for levels a_1 and a_2 indicates why the effects of factor

A are masked by the interaction. For the first two levels of factor B the means for level a_1 are higher than the corresponding means for level a_2 , and for the other two levels the means for level a_1 are lower than the corresponding mean for level a_2 . Opposite algebraic signs of such differences tend to make their sum close to zero in the main effects for factor A .

To illustrate the computation of the simple effects, the variation due to the simple effects of B for level a_2 is obtained as follows (data for the computations are obtained from part ii of Table 6.3-2):

$$(5a_2) = 3.00^2 + 4.25^2 + 10.25^2 + 9.75^2 = 227.19$$

$$(3a_2) = (27.25)^2/4 = 185.64$$

$$SS_{b \text{ for } a_2} = \bar{n}_h[(5a_2) - (3a_2)] = 167.86$$

$$MS_{b \text{ for } a_2} = \frac{SS_{b \text{ for } a_2}}{q - 1} = 55.95.$$

A test of the hypothesis that the variance of the simple effects of factor B at level a_2 is zero is given by the F ratio

$$F = \frac{MS_{b \text{ for } a_2}}{MS_{w. \text{ cell}}} = \frac{55.95}{2.83} = 19.77.$$

The degrees of freedom for this F ratio are (3,25).

In comparing two means, the actual number of observations upon which the mean is based may be used. For example,

$$\begin{aligned} t &= \frac{\overline{AB}_{14} - \overline{AB}_{11}}{\sqrt{MS_{w. \text{ cell}}[(1/n_{14}) + (1/n_{11})]}} \\ &= \frac{9.20 - 5.00}{\sqrt{2.83(\frac{1}{5} + \frac{1}{4})}} = \frac{4.20}{\sqrt{1.27}} = 3.73. \end{aligned}$$

The degrees of freedom for this t statistic are those for $MS_{w. \text{ cell}}$. In making all possible tests between ordered means within a logical grouping, the procedures given in Sec. 6.2 may be followed, assuming \bar{n}_h observations per cell.

6.4 Effect of Scale of Measurement on Interaction

In Sec. 5.4 it was indicated that interactions were in part a function of the choice of the scale of measurement. When this is the case, interaction effects may be

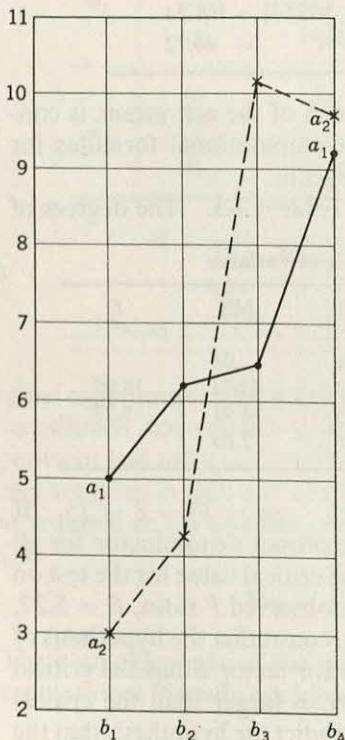


Figure 6.3 Profiles of simple effects for factor B .

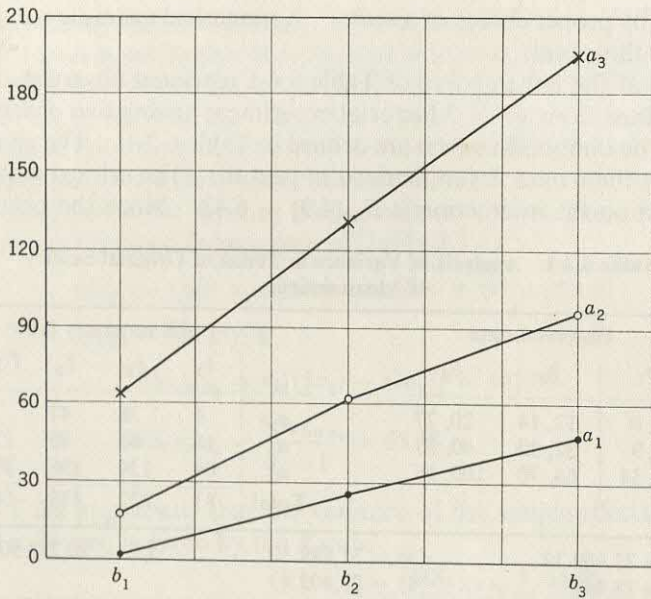


Figure 6.4 Profiles for factor *B* at levels of factor *A*. (Original scale of measurement.)

Table 6.4-2 Analysis of Variance in Terms of Transformed Scale of Measurement

Observed data (transformed scale):

	b_1	b_2	b_3		b_1	b_2	b_3	Total
(i) a_1	1.0, 0.0	3.5, 3.7	4.5, 5.2	a_1	1.0	7.2	9.7	17.9
a_2	3.0, 3.0	5.7, 5.5	6.3, 7.4	a_2	6.0	11.2	13.7	30.9
a_3	5.5, 5.8	8.0, 8.4	10.0, 9.8	a_3	11.3	16.4	19.8	47.5
				Total	18.3	34.8	43.2	96.3

(ii)	(1) = 515.20	(3) = 588.58	(5) = 642.38
	(2) = 643.91	(4) = 568.70	

Source of variation	SS	df	MS	F
A	73.38	2	36.69	$F < 1$
B	53.50	2	26.75	
AB	.30	4	.075	
Within cell	1.53	9	.170	
Total	128.71	17		

In terms of the transformed scale of measurement, the observed data are given in part i of Table 6.4-2. The transformation has the following form.

$$X'_{ijk} = \sqrt{X_{ijk}}.$$

Each entry in the table at the left of part i is the square root of the corresponding entry in part i of Table 6.4-1. The analysis of variance for the transformed data is summarized in part iii of Table 6.4-2. It will be noted that the F ratio in the test on the AB interaction is less than unity in this case. In contrast, the F ratio for the analysis in terms of the original scale

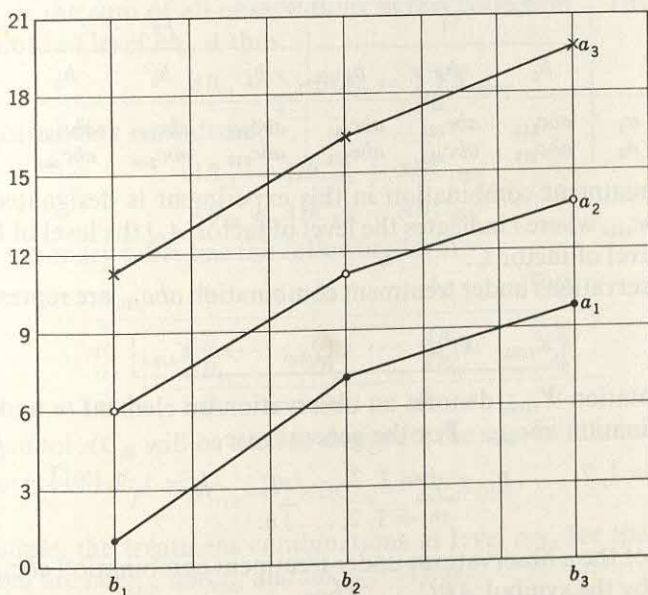


Figure 6.5 Profiles for factor B at levels of factor A . (Transformed scale of measurement.)

was 12.19. The profiles in terms of the transformed scale of measurement are shown in Fig. 6.5. The magnitude of the mean squares for the main effects, relative to the within-cell mean square, is approximately constant for both scales of measurement.

Not all interaction effects can be regarded as functions of the scale of measurement. In cases where profiles cross, or in cases where the profiles have quite different shapes, transformations on the scale of measurement will not remove interaction effects. If, however, interaction effects can be removed by transformations, there are many advantages in working with a model which contains no interaction terms. This is particularly true in the mixed model, which has both fixed and random variables, since interactions between fixed and random factors form denominators for F ratios.

6.5 $p \times q \times r$ Factorial Experiment Having n Observations per Cell

The notation and computational procedures developed in Sec. 6.2 may be extended to three-factor experiments as well as higher-order factorial experiments. In this section the extension will be made to a $p \times q \times r$ factorial experiment. It will be assumed that there are n observations in each cell. Notation will be indicated for the special case of a $2 \times 3 \times 2$ factorial experiment. There are $pqr = 2(3)(2) = 12$ treatment combinations in this experiment. The notation for the treatment combinations is illustrated in the following table:

	c_1			c_2		
	b_1	b_2	b_3	b_1	b_2	b_3
a_1	abc_{111}	abc_{121}	abc_{131}	abc_{112}	abc_{122}	abc_{132}
a_2	abc_{211}	abc_{221}	abc_{231}	abc_{212}	abc_{222}	abc_{232}

A typical treatment combination in this experiment is designated by the notation abc_{ijk} , where i indicates the level of factor A , j the level of factor B , and k the level of factor C .

The n observations under treatment combination abc_{ijk} are represented as follows:

$$X_{ijk1} \quad X_{ijk2} \quad \cdots \quad X_{ijkm} \quad \cdots \quad X_{ijkn}$$

Thus the notation X_{ijkm} denotes an observation on element m under treatment combination abc_{ijk} . For the general case,

$$i = 1, 2, \dots, p; \quad j = 1, 2, \dots, q; \quad k = 1, 2, \dots, r; \\ m = 1, 2, \dots, n.$$

The sum of the n observations under treatment combination abc_{ijk} will be designated by the symbol ABC_{ijk} . Thus,

$$ABC_{ijk} = \sum_m X_{ijkm}.$$

A table of such sums will be called an ABC summary table. For the case of a $2 \times 3 \times 2$ factorial experiment, the ABC summary table has the following form:

	c_1			c_2			Total
	b_1	b_2	b_3	b_1	b_2	b_3	
a_1	ABC_{111}	ABC_{121}	ABC_{131}	ABC_{112}	ABC_{122}	ABC_{132}	A_1
a_2	ABC_{211}	ABC_{221}	ABC_{231}	ABC_{212}	ABC_{222}	ABC_{232}	A_2
Total	BC_{11}	BC_{21}	BC_{31}	BC_{12}	BC_{22}	BC_{32}	G

The column totals in this ABC summary table have the general form

$$BC_{jk} = \sum_i ABC_{ijk} = \sum_i \sum_m X_{ijkm}.$$

That is, a column total represents the sum of all observations under treatment combination bc_{jk} , the levels of factor A being disregarded. The BC summary table has the following form:

	b_1	b_2	b_3	Total
c_1	BC_{11}	BC_{21}	BC_{31}	C_1
c_2	BC_{12}	BC_{22}	BC_{32}	C_2
Total	B_1	B_2	B_3	G

Treatment combination ab_{ij} is defined to be the collection of treatment combinations $abc_{ij1}, abc_{ij2}, \dots, abc_{ijr}$. The sum of all observations at level ab_{ij} is the sum of all observations in this collection. The sum of all observations at level ab_{ij} is thus

$$AB_{ij} = \sum_k ABC_{ijk} = \sum_k \sum_m X_{ijkm}.$$

For the case being considered,

$$AB_{11} = ABC_{111} + ABC_{112},$$

$$AB_{23} = ABC_{231} + ABC_{232}.$$

The AB summary table has the following form:

	b_1	b_2	b_3	Total
a_1	AB_{11}	AB_{12}	AB_{13}	A_1
a_2	AB_{21}	AB_{22}	AB_{23}	A_2
Total	B_1	B_2	B_3	G

The symbol AC_{ik} will be used to designate the sum of all observations at level ac_{ik} . Thus,

$$AC_{ik} = \sum_j ABC_{ijk} = \sum_j \sum_m X_{ijkm}.$$

For example, the treatment combinations at level ac_{12} for the case being considered are abc_{112}, abc_{122} , and abc_{132} . Thus,

$$AC_{12} = ABC_{112} + ABC_{122} + ABC_{132}.$$

The AC summary table has the following form:

	c_1	c_2	Total
a_1	AC_{11}	AC_{12}	A_1
a_2	AC_{21}	AC_{22}	A_2
Total	C_1	C_2	G_3

The sum of all observations at level a_1 may be obtained as follows:

$$A_i = \sum_j AB_{ij} = \sum_k AC_{ik} = \sum_j \sum_k ABC_{ijk} = \sum_j \sum_k \sum_m X_{ijkm}.$$

The sum of all observations at level b_j is given by

$$B_j = \sum_i AB_{ij} = \sum_k BC_{jk} = \sum_i \sum_k ABC_{ijk} = \sum_i \sum_k \sum_m X_{ijkm}.$$

Similarly,

$$C_k = \sum_i AC_{ik} = \sum_j BC_{jk} = \sum_i \sum_j ABC_{ijk} = \sum_i \sum_j \sum_m X_{ijkm}.$$

To summarize, two-way summary tables are most readily obtained from three-way summary tables by combining levels of one of the factors. Thus the BC summary table is obtained from the ABC summary table by adding totals in the latter table which are at the same levels of factors B and C but at different levels of factor A . The AB summary table is obtained from the ABC summary table by adding totals in the latter table which are at the same levels of factors A and B but at different levels of factor C . The AC summary table is obtained in an analogous manner.

Table 6.5-1 Definition of Computational Symbols

(i)

$$\begin{aligned}
 (1) &= G^2/npqr & (6) &= [\Sigma(AB_{ij})^2]/nr \\
 (2) &= \Sigma X_{ijk}^2 & (7) &= [\Sigma(AC_{ik})^2]/nq \\
 (3) &= (\Sigma A_i^2)/nqr & (8) &= [\Sigma(BC_{jk})^2]/np \\
 (4) &= (\Sigma B_j^2)/npr & (9) &= [\Sigma(ABC_{ijk})^2]/n \\
 (5) &= (\Sigma C_k^2)/npq
 \end{aligned}$$

(ii)

$$\begin{aligned}
 SS_a &= nqr\Sigma(\bar{A}_i - \bar{G})^2 = (3) - (1) \\
 SS_b &= npr\Sigma(\bar{B}_j - \bar{G})^2 = (4) - (1) \\
 SS_c &= npq\Sigma(\bar{C}_k - \bar{G})^2 = (5) - (1) \\
 SS_{ab} &= nr\Sigma(\bar{AB}_{ij} - \bar{A}_i - \bar{B}_j + \bar{G})^2 \\
 &= nr\Sigma(\bar{AB}_{ij} - \bar{G})^2 - SS_a - SS_b = (6) - (3) - (4) + (1) \\
 SS_{ac} &= nq\Sigma(\bar{AC}_{ik} - \bar{A}_i - \bar{C}_k + \bar{G})^2 \\
 &= nq\Sigma(\bar{AC}_{ik} - \bar{G})^2 - SS_a - SS_c = (7) - (3) - (5) + (1) \\
 SS_{bc} &= np\Sigma(\bar{BC}_{jk} - \bar{B}_j - \bar{C}_k + \bar{G})^2 \\
 &= np\Sigma(\bar{BC}_{jk} - \bar{G})^2 - SS_b - SS_c = (8) - (4) - (5) + (1) \\
 SS_{abc} &= n\Sigma(\bar{ABC}_{ijk} - \bar{AB}_{ij} - \bar{AC}_{ik} - \bar{BC}_{jk} + \bar{A}_i + \bar{B}_j + \bar{C}_k - \bar{G})^2 \\
 &= n\Sigma(\bar{ABC}_{ijk} - \bar{G})^2 - SS_{ab} - SS_{ac} - SS_{bc} - SS_a - SS_b - SS_c \\
 &= (9) - (6) - (7) - (8) + (3) + (4) + (5) - (1) \\
 SS_{w, cell} &= \Sigma(X_{ijk} - \bar{ABC}_{ijk})^2 = (2) - (9) \\
 SS_{total} &= \Sigma(X_{ijk} - \bar{G})^2 = (2) - (1)
 \end{aligned}$$

Symbols in terms of which computational formulas may be conveniently written are summarized in part i of Table 6.5-1. With the exception of symbol (2), all are obtained from either the two-way or the three-way summary tables. In each case the divisor for a computational symbol is the number of basic observations summed to obtain a term which is squared in the numerator. For example, in computational symbol (3), A_i is squared. There are nqr observations summed to obtain A_i .

There is a relatively simple method of determining the number of basic observations summed to obtain a total of the form A_i . In a $p \times q \times r$ factorial experiment a basic observation is represented by the symbol X_{ijk} . In A_i the subscripts j , k , and m are missing. The numbers of levels corresponding to these missing subscripts are, respectively, q , r , and n . The number of observations summed to obtain A_i is the product of these missing

subscripts, qnr . In the total B_j , the subscripts i, k , and m are missing. The corresponding numbers of levels are p, r , and n ; hence the number of observations summed is prn . In the total BC_{jk} the subscripts i and m are missing. Hence pn observations are summed in this total. In the total AC_{ik} the subscripts j and m are missing. Hence the number of observations summed in this total is qn .

Where the index of summation does not appear under the summation symbol, it is understood that the summation is over all possible terms of the form specified. For example, the notation $\Sigma(AB_{ij})^2$ indicates that the sum is over all the pq cell totals in a two-way summary table of the form AB_{ij} . Similarly, the notation $\Sigma(BC_{jk})^2$ indicates that the sum is over all the qr cell totals having the form BC_{jk} . The basic definitions of the estimates of variation due to main effects and interactions are summarized in part ii of Table 6.5-1. Corresponding computational formulas are also given.

In a three-factor factorial experiment there are various orders of simple effects. Computational formulas for these sources of variation may be obtained by specializing the symbols given in part i of Table 6.5-1. The symbol $(6a_i)$ is defined to be the equivalent of symbol (6), in which the summation is restricted to level a_i . For example,

$$(6a_1) = \frac{(AB_{11})^2 + (AB_{12})^2 + \cdots + (AB_{1q})^2}{nr}$$

$$(6a_2) = \frac{(AB_{21})^2 + (AB_{22})^2 + \cdots + (AB_{2q})^2}{nr}$$

Similarly the symbol $(7c_k)$ is defined to be computational symbol (7), in which the summation is limited to level c_k —that is, the summation is restricted to row c_k of the BC summary table.

The computational symbol $(9a_i)$ restricts the summation in (9) to row a_i of the ABC summary table. Computational symbol $(9ab_{ij})$ restricts the summation in (9) to just those totals in which the factor A is at level a_i and factor B is at level b_j . For example,

$$(9ab_{12}) = \frac{(ABC_{121})^2 + (ABC_{122})^2 + \cdots + (ABC_{12r})^2}{n}$$

$$(9ab_{23}) = \frac{(ABC_{231})^2 + (ABC_{232})^2 + \cdots + (ABC_{23r})^2}{n}$$

In terms of computational symbols defined in this manner, computational formulas for various orders of simple effects are given in Table 6.5-2.

The following relationships hold for the computational symbols:

$$(6) = (6a_1) + (6a_2) + \cdots + (6a_p),$$

$$(7) = (7a_1) + (7a_2) + \cdots + (7a_p),$$

$$(9) = (9a_1) + (9a_2) + \cdots + (9a_p),$$

$$(9) = \sum_i \sum_j (9ab_{ij}).$$

Analogous relations hold for other computational symbols.

Numerical Example. The computational procedures will be illustrated by means of a $2 \times 3 \times 2$ factorial experiment. The purpose of this experiment is to evaluate the relative effectiveness of three methods of training

Table 6.5-2 Computational Formulas for Simple Effects

Effects	Sum of squares
Simple interactions:	
AB for level c_k	$(9c_k) - (7c_k) - (8c_k) + (5c_k)$
AC for level b_j	$(9b_j) - (6b_j) - (8b_j) + (4b_j)$
BC for level a_i	$(9a_i) - (6a_i) - (7a_i) + (3a_i)$
Simple main effects:	
A for level b_j	$(6b_j) - (4b_j)$
A for level c_k	$(7c_k) - (5c_k)$
B for level a_i	$(6a_i) - (3a_i)$
B for level c_k	$(8c_k) - (5c_k)$
C for level a_i	$(7a_i) - (3a_i)$
C for level b_j	$(8b_j) - (4b_j)$
Simple, simple main effects:	
A for level bc_{jk}	$(9bc_{jk}) - (8bc_{jk})$
B for level ac_{ik}	$(9ac_{ik}) - (7ac_{ik})$
C for level ab_{ij}	$(9ab_{ij}) - (6ab_{ij})$
Computational checks:	
$\sum_k SS_a \text{ for } c_k = SS_a + SS_{ac}$	
$\sum_k SS_{ab} \text{ for } c_k = SS_{ab} + SS_{abc}$	
$\sum_{j,k} SS_a \text{ for } bc_{jk} = SS_a + SS_{ab} + SS_{ac} + SS_{abc}$	

(factor B). Two instructors (factor C) are used in the experiment; subjects in the experiment are classified on the basis of educational background (factor A). The plan for this experiment may be represented as follows:

Instructor:		c_1			c_2		
Training method:		b_1	b_2	b_3	b_1	b_2	b_3
Educational level	a_1	G_{111}	G_{121}	G_{131}	G_{112}	G_{122}	G_{132}
	a_2	G_{211}	G_{221}	G_{231}	G_{212}	G_{222}	G_{232}

In this plan G_{111} represents a group of subjects at educational level a_1 assigned to instructor c_1 to be trained under method b_1 . The symbol G_{132} denotes the group of subjects at educational level a_1 assigned to instructor c_2

to be trained under method b_3 . Thus each instructor teaches groups from both educational levels under each of the training methods. It will be assumed that there are 10 subjects in each of the groups, a total of 120 subjects in all.

In this experiment the methods of training (factor B) and the levels of education (factor A) will be considered fixed factors. Factor A is a classification variable included in the experiment to control potential variability in the experimental units, which is a function of level of education. (All relevant levels of the education factor must be covered if this factor is fixed.) Factor B is the treatment variable of primary interest; this factor is directly

Table 6.5-3 Expected Values of Mean Squares for Numerical Example
(A, B fixed; C random)

Effect	i	j	k	m	Expected value of mean square
α_i	0	3	2	10	$\sigma_e^2 + 30\sigma_{\alpha\gamma}^2 + 60\sigma_\alpha^2$
β_j	2	0	2	10	$\sigma_e^2 + 20\sigma_{\beta\gamma}^2 + 40\sigma_\beta^2$
γ_k	2	3	1	10	$\sigma_e^2 + 60\sigma_\gamma^2$
$\alpha\beta_{ij}$	0	0	2	10	$\sigma_e^2 + 20\sigma_{\alpha\beta}^2$
$\alpha\gamma_{ik}$	0	3	1	10	$\sigma_e^2 + 30\sigma_{\alpha\gamma}^2$
$\beta\gamma_{jk}$	2	0	1	10	$\sigma_e^2 + 20\sigma_{\beta\gamma}^2$
$\alpha\beta\gamma_{ijk}$	0	0	1	10	$\sigma_e^2 + 10\sigma_{\alpha\beta\gamma}^2$
$\varepsilon_{m(ijk)}$	1	1	1	1	σ_e^2

under the control of the experimenter. There is some question about whether or not factor C should be considered a fixed variable. If it is the purpose of the experiment to draw inferences about the methods of training which potentially hold for a population of instructors, of which the two instructors in the experiment can be considered a random sample, then the instructor factor is random. If inferences about the methods are to be limited to the two instructors in the experiment, then the instructor factor is fixed. Often in this type of experiment, inferences are desired about the methods over a population of specified instructors. Hence the instructor factor should be considered as a random factor, and suitable randomization procedures are required in the selection of the instructors.

The expected values for the mean squares for the case under consideration are given in Table 6.5-3. The model from which these expected values were obtained includes interaction terms with the instructor factor. According to these expected values, the test on the main effect of factor B has the form

$$F = \frac{MS_b}{MS_{be}}.$$

This F ratio has degrees of freedom $[(q - 1), (q - 1)(r - 1)]$, which in this case is (2,2). When the denominator of an F ratio has only two degrees of

Suppose that the ABC summary table for the data obtained in the experiment is that given in part i of Table 6.5-4. Each of the entries in this table is the sum of the 10 criterion scores for the corresponding group of subjects. For example, the entry in cell abc_{132} is the sum of the 10 criterion scores for the subjects in group G_{132} , that is, the subjects at education level a_1 , trained under method b_3 by instructor c_2 . The two-way summary tables given in part i are obtained from the three-way summary table. For example, the entry ab_{11} in the AB summary table is given by

$$\begin{aligned} AB_{11} &= ABC_{111} + ABC_{112} \\ &= 20 + 16 = 36. \end{aligned}$$

The computational symbols defined in part i of Table 6.5-1 are obtained

Table 6.5-5 Summary of Analysis of Variance

Source of variation	SS	df	MS
A (level of education)	122.01	$p - 1 = 1$	122.01
B (methods of training)	25.31	$q - 1 = 2$	12.66
C (instructors)	0.41	$r - 1 = 1$	0.41
AB	23.12	$(p - 1)(q - 1) = 2$	11.56
AC	2.40	$(p - 1)(r - 1) = 1$	2.40
BC	1.72	$(q - 1)(r - 1) = 2$	0.86
ABC	0.32	$(p - 1)(q - 1)(r - 1) = 2$	0.16
Within cell (experimental error)	110.70	$pqr(n - 1) = 108$	1.02
Total	285.99	119	

in part ii in Table 6.5-4. Data for all these computations, with the exception of symbol (2), are contained in part i. Symbol (2) is obtained from the individual criterion scores; the latter are not given in this table. The computation of the sums of squares is completed in part iii.

The analysis of variance is summarized in Table 6.5-5. Preliminary tests on the model will be made on the interactions with factor C (instructors) before proceeding with other tests. According to the expected values of the mean squares given in Table 6.5-3, tests on interactions with factor C all have $MS_{w, cell}$ as a denominator. These tests have the following form:

$$F = \frac{MS_{abc}}{MS_{w, cell}} = \frac{0.16}{1.02} = 0.15,$$

$$F = \frac{MS_{ac}}{MS_{w, cell}} = \frac{2.40}{1.02} = 2.33,$$

$$F = \frac{MS_{bc}}{MS_{w, cell}} = \frac{0.86}{1.02} = 0.85.$$

Only the F ratio for the test on the AC interaction is greater than unity. By use of the .10 level of significance, the critical value for the latter test is $F_{.90}(2, 108) = 2.76$.

The outcome of these preliminary tests on the model does not contradict the hypothesis that interactions with factor C may be considered negligible. (The AC interaction is a borderline case.) On a priori grounds, if the instructors are carefully trained, interaction effects with instructors may often be kept relatively small. On these bases, the decision is made to drop interactions with factor C from the model. The expected values corresponding to this revised model are obtained from Table 6.4-3 by dropping the terms $\sigma_{\alpha\gamma}^2$, $\sigma_{\beta\gamma}^2$, and $\sigma_{\alpha\beta\gamma}^2$ from the expected values in this table. Since the degrees of freedom for the within-cell variation (108) are large relative to the degrees of freedom for the interactions with factor C , which total 5, pooling these interactions with the within-cell variation will not appreciably affect the magnitude of $MS_{w. cell}$ or the degrees of freedom for the resulting pooled error term. Hence $MS_{w. cell}$ is used under the revised model for final tests on factors A and B . These tests have the following form:

$$F = \frac{MS_{ab}}{MS_{w. cell}} = 11.33, \quad F_{.99}(2, 108) = 4.82;$$

$$F = \frac{MS_b}{MS_{w. cell}} = 12.41, \quad F_{.99}(2, 108) = 4.82;$$

$$F = \frac{MS_a}{MS_{w. cell}} = 119.62, \quad F_{.99}(1, 108) = 6.90.$$

Because of the significant AB interaction, care must be taken in interpreting the main effects due to factor B . The manner in which educational level is related to method of training is most readily shown by the profiles of the simple effects for the methods at each of the levels of education. These profiles are drawn in Fig. 6.6. Data for these profiles are obtained from the AB summary table. Inspection of these profiles indicates that differences between the methods of training for groups at level a_2 are not so marked as the corresponding differences for groups at level a_1 .

Variation due to differences between training methods for groups at level a_1 is given by (data are obtained from row a_1 of the AB summary table)

$$SS_{b \text{ at } a_1} = \frac{36^2 + 63^2 + 20^2}{20} - \frac{119^2}{60} = 47.23.$$

This source of variation has $q - 1 = 2$ degrees of freedom. Hence the mean square of the simple effect of factor B for level a_1 is

$$MS_{b \text{ for } a_1} = \frac{47.23}{2} = 23.62.$$

To test the hypothesis of no difference between the methods of training for groups at educational level a_1 ,

$$F = \frac{MS_{b \text{ for } a_1}}{MS_{w. cell}} = \frac{23.62}{1.02} = 23.15.$$

The degrees of freedom for this F ratio are (2,108). The data clearly indicate a significant difference between the methods for level a_1 . The test on the simple effects of methods for level a_2 indicates no significant difference between the methods; the F for this test will be found to be $F = 0.60/1.02$.

The levels of factor C (instructors) in this design may be considered to be replications. As indicated by the expected values for the mean squares, the

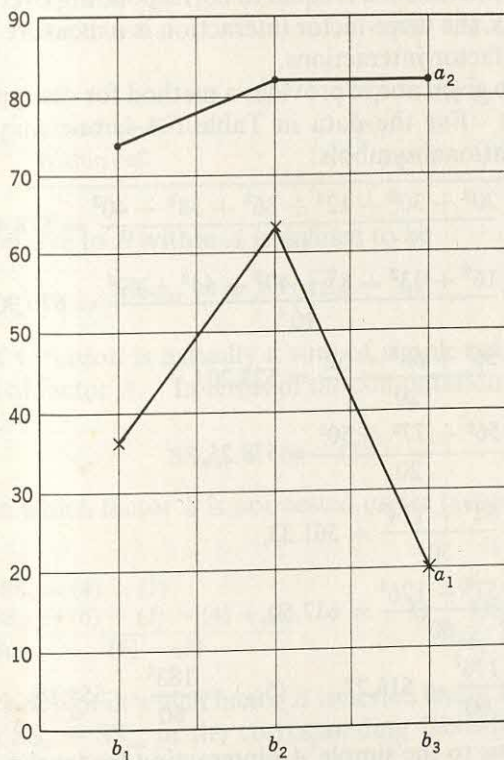


Figure 6.6 Profiles of simple main effects for training methods.

proper denominator for tests on fixed effects is the corresponding interaction with replications, provided that such interactions cannot be pooled. When tests are made by using a pooled error term (this is essentially what has been done in the example that has just been considered), it is implicitly assumed that interactions with replications (instructors) do not exist.

There are potentially many reasons for the presence of interaction in this kind of experiment. Inspection of the $A \times B$ summary table indicates that groups at level a_2 are uniformly good under all the training methods. In part, this may be a function of a ceiling effect on the criterion. If the latter is the case, then the interaction is an artifact of the way in which performance is measured. Care is required in constructing the criterion to avoid this kind of artifact.

Alternative Computational Procedures for Three-factor Interactions. In Sec. 5.10 the relationship between two-factor and three-factor interactions was indicated. For a $2 \times 3 \times 2$ factorial experiment,

$$SS_{abc} = SS_{ab \text{ at } c_1} + SS_{ab \text{ at } c_2} - SS_{ab}.$$

In general, a three-factor interaction will be zero whenever the sum of the simple two-factor interactions is equal to corresponding over-all two-factor interaction. Thus, the three-factor interaction is a measure of the additivity of simple two-factor interactions.

The relationship given above provides a method for computing the three-factor interaction. For the data in Table 6.4-4, one may compute the following computational symbols:

$$(9c_1) = \frac{20^2 + 30^2 + 12^2 + 36^2 + 38^2 + 40^2}{10} = 578.40,$$

$$(9c_2) = \frac{16^2 + 33^2 + 8^2 + 40^2 + 44^2 + 42^2}{10} = 670.90,$$

$$(8c_1) = \frac{56^2 + 68^2 + 52^2}{20} = 523.20,$$

$$(8c_2) = \frac{56^2 + 77^2 + 50^2}{20} = 578.25,$$

$$(7c_1) = \frac{62^2 + 114^2}{30} = 561.33,$$

$$(7c_2) = \frac{57^2 + 126^2}{30} = 637.50,$$

$$(5c_1) = \frac{176^2}{60} = 516.27, \quad (5c_2) = \frac{183^2}{60} = 558.15.$$

The variation due to the simple AB interaction for level c_1 is

$$(9c_1) - (7c_1) - (8c_1) + (5c_1) = 10.14.$$

The corresponding variation for level c_2 is

$$(9c_2) - (7c_2) - (8c_2) + (5c_2) = 13.30.$$

The variation due to the over-all AB interaction was found to be 23.12. Hence the variation due to ABC is

$$SS_{abc} = 10.14 + 13.30 - 23.12 = 0.32.$$

6.6 Computational Procedures for Nested Factors

Factorial designs in which one or more factors are nested were discussed in Sec. 5.12. Computational procedures in which one factor is nested under a second will be illustrated from the case of a $p \times q \times r$ factorial

experiment having n observations in each cell. Assume that factor B is nested under factor A . The analysis of variance for this case generally takes the following form:

Source of variation	df
A	$p - 1$
B within A	$p(q - 1)$
C	$r - 1$
AC	$(p - 1)(q - 1)$
$(B \text{ within } A) \times C$	$p(q - 1)(r - 1)$
Within cell	$pqr(n - 1)$

The variation due to B within A is defined to be

$$SS_{b(a)} = nr \sum_i \sum_j (\overline{AB}_{ij} - \bar{A}_i)^2.$$

This source of variation is actually a sum of simple main effects of factor B at each level of factor A . In terms of the computational symbols defined in Table 6.5-1,

$$SS_{b(a)} = (6) - (3).$$

For a design in which factor B is not nested under factor A ,

	df
$SS_b = (4) - (1)$	$q - 1$
$SS_{ab} = (6) - (3) - (4) + (1)$	$(p - 1)(q - 1)$
Sum $(6) - (3)$	$p(q - 1)$

Thus $SS_{b(a)}$ in a design in which factor B is nested under factor A is numerically equal to $SS_b + SS_{ab}$ in the corresponding factorial design in which factor B is not nested under factor A .

The variation due to the interaction $(B \text{ within } A) \times C$ is defined to be

$$SS_{b(a)c} = n \sum_i \sum_j \sum_k (\overline{ABC}_{ijk} - \overline{AB}_{ij} - \overline{AC}_{ik} - \bar{A}_i)^2$$

This source of variation is actually a sum of simple interactions. In terms of the computational symbols defined in Table 6.5-1,

$$SS_{b(a)c} = (9) - (6) - (7) + (3).$$

For a $p \times q \times r$ factorial experiment in which there is no nested factor,

	df
$SS_{bc} = (8) - (4) - (5) + (1)$	$(q - 1)(r - 1)$
$SS_{abc} = (9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)$	$(p - 1)(q - 1)(r - 1)$
Sum $(9) - (6) - (7) + (3)$	$p(q - 1)(r - 1)$

Suppose that the purpose of this experiment is to test the relative effectiveness of two drugs on patients in different diagnostic categories. Patients from hospitals $b_{1(a_1)}$ and $b_{2(a_1)}$ are given drug a_1 ; patients from hospitals $b_{1(a_2)}$ and $b_{2(a_2)}$ are given drug a_2 . Since only one of the drugs under study is administered within a hospital, the hospital factor is nested under the drug factor. The diagnostic categories are considered to be comparable across

Table 6.6-2 Numerical Example

ABC summary table:

	a_1		a_2		Total
	$b_{1(a_1)}$	$b_{2(a_1)}$	$b_{1(a_2)}$	$b_{2(a_2)}$	
c_1	15	18	30	35	98
c_2	22	25	24	21	92
c_3	38	41	10	14	103
Total	75	84	64	70	293

(i) *AC* summary table:

	a_1	a_2	Total
c_1	33	65	98
c_2	47	45	92
c_3	79	24	103
Total	159	134	293

	(1) = $(293)^2/60$	= 1430.82
	(2) (not obtained from above summary tables)	= 1690
	(3) = $(159^2 + 134^2)/30$	= 1441.23
	(5) = $(98^2 + 92^2 + 103^2)/20$	= 1433.85
	(6) = $(75^2 + 84^2 + 64^2 + 70^2)/15$	= 1445.13
	(7) = $(33^2 + 47^2 + \dots + 45^2 + 24^2)/10$	= 1636.50
	(9) = $(15^2 + 22^2 + \dots + 21^2 + 14^2)/5$	= 1644.20

all hospitals; hence factor C is not nested. Only a random sample of the population of hospitals about which inferences are to be drawn is included in the experiment; hence factor B is random.

Suppose that the *AB* summary table for the data obtained in this experiment is that in part i of Table 6.6-2. With the exception of symbol (2), which is computed from the individual observations, data for the computations are given in part i. (The definitions of these symbols appear in Table 6.5-1.) Symbols (4) and (8) are not required for this case. The analysis of variance is summarized in Table 6.6-3.

The expected values of the mean squares are given in Table 6.6-1. Before testing the main effects and interaction of the fixed factors, preliminary tests on the model are made with respect to factor B and its interaction with

factor C . These preliminary tests will be made at the .10 level of significance. Inspection of the expected values of the mean squares indicates that the preliminary tests have the following form.

$$F = \frac{MS_{b(a)}}{MS_{w, cell}} = 1.37, \quad F_{.90}(3, 48) = 2.22;$$

$$F = \frac{MS_{b(a)c}}{MS_{w, cell}} < 1.00, \quad F_{.90}(6, 48) = 1.92.$$

Table 6.6-3 Summary of Analysis of Variance

(i)

$$\begin{aligned}
 SS_a &= (3) - (1) &= 10.41 \\
 SS_{b(a)} &= (6) - (3) &= 3.90 \\
 SS_c &= (5) - (1) &= 3.03 \\
 SS_{ac} &= (7) - (3) - (5) + (1) &= 192.24 \\
 SS_{b(a)c} &= (9) - (6) - (7) + (3) &= 3.80 \\
 SS_{w, \text{ cell}} &= (2) - (9) &= 45.80
 \end{aligned}$$

(ii)

Source of variation	SS	df	MS
<i>A</i> (drugs)	10.41	1	10.41
<i>B</i> (hospitals within <i>A</i>)	3.90	3	1.30
<i>C</i> (categories)	3.03	2	1.52
<i>AC</i>	192.24	2	96.12
<i>B(A)</i> × <i>C</i>	3.80	6	0.63
Within cell	45.80	48	0.95
Pooled error	53.50	57	0.94

Neither of the F ratios exceeds specified critical values. Hence variation due to $B(A)$ and $B(A) \times C$ is pooled with the within-cell variation. Thus,

$$SS_{\text{pooled error}} = SS_{b(a)} + SS_{b(a)c} + SS_{w, cell}.$$

The degrees of freedom for this term are the sum of the respective degrees of freedom for the parts.

The denominator for all final tests is $MS_{\text{pooled error}}$. For this case the final tests are

$$F = \frac{MS_{ac}}{MS_{\text{pooled error}}} = 102.26, \quad F_{.99}(2, 57) = 5.00;$$

$$F = \frac{MS_a}{MS_{\text{pooled error}}} = 11.07, \quad F_{.99}(1, 57) = 7.10;$$

$$F = \frac{MS_c}{MS_{\text{pooled error}}} = 1.62, \quad F_{.99}(2, 57) = 5.00.$$

In spite of the significant AC interaction, the main effect for factor A is significant. An analysis of the simple effects is required for an adequate interpretation of the effects of the drugs. Inspection of the AC summary

table indicates that drug a_2 has the higher criterion total for category c_1 ; there appears to be little difference between the criterion scores for category c_2 ; drug a_1 has the higher criterion total for category c_3 . Formal tests on these last statements have the following form:

$$F = \frac{(AC_{11} - AC_{21})^2}{2nqMS_{\text{pooled error}}} = \frac{(33 - 65)^2}{2(5)(2)(0.94)} = 54.47,$$

$$F = \frac{(AC_{12} - AC_{22})^2}{2nqMS_{\text{pooled error}}} = \frac{(47 - 45)^2}{2(5)(2)(0.94)} = 0.21,$$

$$F = \frac{(AC_{13} - AC_{23})^2}{2nqMS_{\text{pooled error}}} = \frac{(79 - 24)^2}{2(5)(2)(0.94)} = 160.90.$$

6.7 Factorial Experiment with a Single Control Group

A design closely related to one reported by Levison and Zeigler (1959) will be used to illustrate the material that will be discussed in this section. The purpose of this experiment is to test the effect of amount and time of irradiation upon subsequent learning ability. Different dosages (factor B) of irradiation are administered at different age levels (factor A). When all subjects reach a specified age, they are given a series of learning tasks. Separate analyses are made of the criterion scores for each task. The plan for the experiment may be represented as follows:

		Dosage of irradiation	
		b_1	b_2
Age at which irradiation is administered	a_1	G_{11}	G_{12}
	a_2	G_{21}	G_{22}
	a_3	G_{31}	G_{32}
	a_4	G_{41}	G_{42}

Suppose that each of the above groups contains 10 subjects. In addition to these eight groups, there is a control group G_0 , having 20 subjects, which receives no irradiation treatment.

Thus there are nine groups in all, eight groups in the cells of the 4×2 factorial experiment plus a control group. For the general case there will be $pq + 1$ groups. The analysis of variance for this experimental plan may take the following form:

Between groups	$(pq + 1) - 1 = 8$
Control vs. all others	1
A (age)	$p - 1 = 3$
B (dosage)	$q - 1 = 1$
AB	$(p - 1)(q - 1) = 3$
Within cell	$pq(n - 1) + (n_0 - 1) = 91$

The one degree of freedom for control vs. all others contrasts the control group with all the experimental groups combined. It is also of interest to contrast the control group with each experimental group. The procedure described in Sec. 3.10 may be adapted for this purpose. In this case the t statistic has the form

$$t = \frac{\bar{C}_0 - \overline{AB}_{ij}}{\sqrt{MS_{w, cell}[(1/n_0) + (1/n)]}},$$

where \bar{C}_0 represents the mean of the control group. Critical values for this t statistic are given in Table B.6. The degrees of freedom for this statistic are those of $MS_{w, cell}$; k corresponds to the total number of groups, which in this case is $pq + 1$.

In the analysis of variance for the factorial part of the experiment the control group is disregarded. An alternative analysis considers the control condition as defining an added level for one factor. That is, the plan is considered to be a $p \times (q + 1)$ factorial experiment. In the latter case, the plan may be represented as follows:

		Dosage of irradiation		
		b_0	b_1	b_2
Age at which irradiation is administered	a_1		G_{11}	G_{12}
	a_2	G_0	G_{21}	G_{22}
	a_3		G_{31}	G_{32}
	a_4		G_{41}	G_{42}

Since group G_0 received no irradiation, it cannot be classified along the age dimension. From one point of view, however, G_0 may be considered as belonging to all the age classifications. The author does not agree, in general, with this latter point of view, but if it makes good experimental sense to adopt this point of view, then the data from group G_0 are used in all cells under b_0 . If this is done, the analysis of variance has the following form:

A (age)		$p - 1 = 3$
B (dosage)		$q = 2$
AB		$p(q - 1) = 6$
Within cell	$pq(n - 1) + (n_0 - 1) - (p - 1) = 88$	

Since the data from G_0 are used in p different cells of the experiment, the data within each cell are not statistically independent. To compensate, in part, for this lack of statistical independence, $p - 1$ degrees of freedom are subtracted from the estimate of experimental error.

Numerical Example. Suppose data obtained from the eight experimental groups in the experiment which has just been described are those given in part i of Table 6.7-1. Each entry in the AB summary table is the

sum of 10 observations. The computational symbols in part i are defined in Table 6.5-1. [Data for the computation of symbol (2) are not given.] Summary data for the control group are given in part ii. It is noted that the number of subjects in the control group is 20, whereas the number of subjects in each of the experimental groups is 10.

Table 6.7-1 Numerical Example

AB summary table ($n = 10$):

	b_1	b_2	Total	
(i) a_1	380	310	690	(1) = 143,143
a_2	405	340	745	(2) = 148,129
a_3	485	470	955	(3) = 146,559
a_4	504	490	994	(4) = 143,479
	1774	1610	3384	(5) = 147,037

Data for control group ($n_0 = 20$):

$$(ii) \quad C_0 = \Sigma X = 1000 \quad \Sigma X^2 = 50,300 \quad SS_0 = 50,300 - \frac{(1000)^2}{20} = 300$$

$$(iii) \quad SS_{\text{control vs. all others}} = \frac{[\frac{1}{2}(1000) - (3384)/8]^2}{10[1 + (\frac{1}{8})]} = 527$$

Source of variation	SS	df	MS	F
Control vs. all others	527	1	527	34.44
A (age)	3416	3	1139	
(iv) B (dosage)	336	1	336	
AB	142	3	47.3	3.09
Within cell	$1092 + 300 = 1392$	91	15.30	

The comparison between the control group and all other groups combined has the general form (assuming each mean is based upon n observations)

$$SS_{\text{control vs. all others}} = \frac{n[pqC_0 - \Sigma \Sigma AB_{ij}]^2}{p^2q^2 + pq} = \frac{[(n/n_0)C_0 - (1/pq)G]^2}{n[1 + (1/pq)]}$$

The numerical value of this expression is obtained in part iii. A summary of the analysis of variance is given in part iv. The within-cell variation is the pooled variation for the eight experimental groups and the control group. The degrees of freedom for the within-cell variation are

$$pq(n - 1) + (n_0 - 1) = 8(9) + 19 = 91.$$

The test on the interaction indicates that this source of variation is significantly greater than zero, $F_{.95}(3,91) = 2.71$. The profiles for factor A

at levels b_1 and b_2 as well as the control, b_0 , are plotted in Fig. 6.7. (Data for the control group are in units which are comparable with those in the experimental groups.) Inspection of these profiles indicates relatively large differences between the control group and the groups which were irradiated at ages a_1 and a_2 . There is relatively little difference between the

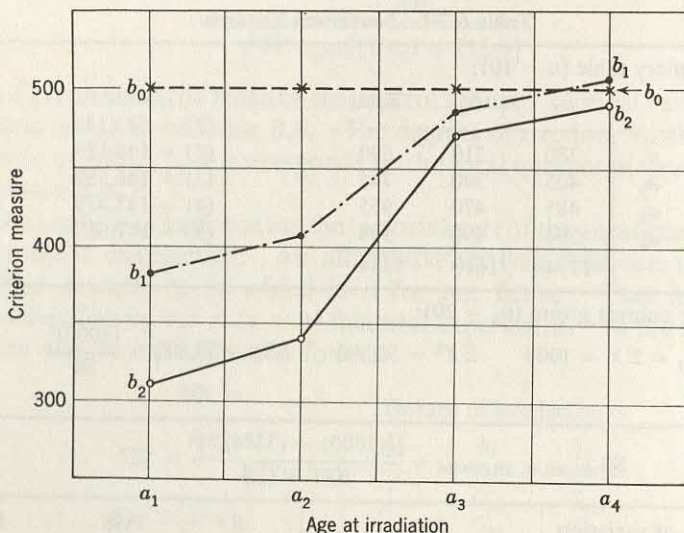


Figure 6.7 Profiles for different dosages.

control group and the groups given irradiation at ages a_3 and a_4 . Further, for groups given irradiation at ages a_1 and a_2 , the groups given the larger dosage (b_2) showed the greater decrement in performance. A formal test on this latter statement uses the statistic

$$\begin{aligned}
 F &= \frac{[(AB_{11} + AB_{21}) - (AB_{12} + AB_{22})]^2}{4nMS_{w. cell}} \\
 &= \frac{[(380 + 405) - (310 + 340)]^2}{40(15.30)} = 29.78.
 \end{aligned}$$

The critical value for this test is $F_{.95}(1,91) = 3.95$.

To test the hypothesis that groups given irradiation at ages a_3 and a_4 do not differ from the control group, the following comparison may be used:

$$\begin{aligned}
 F &= \frac{\left[\frac{2C_0}{n_0/nq} - (A_3 + A_4) \right]^2}{6nqMS_{w. cell}} \\
 &= \frac{\left[\frac{2(1000)}{20/20} - (955 + 994) \right]^2}{120(15.30)} = 1.42.
 \end{aligned}$$

For a test at the .05 level of significance, the critical value for this statistic is $F_{.95}(1,91) = 3.95$. Thus the data indicate there is no statistically significant difference between the groups irradiated at ages a_3 and a_4 and the control group with respect to the criterion of performance measured.

The alternative analysis, which considers the control group as a part of the factorial experiment, is summarized in Table 6.7-2. The entries in

Table 6.7-2 Numerical Example

AB summary table:					
	b_0	b_1	b_3	Total	
(i)	a_1	500	380	310	1190
	a_2	500	405	340	1245
	a_3	500	485	470	1455
	a_4	500	504	490	1494
		2000	1774	1610	5384
	Source of variation	SS	df	MS	F
(ii)	A (age)	2278	3	759	13.46
	B (dosage)	1917	2	986	
	AB	1280	6	213	
	Within cell	1392	88	15.82	

column b_0 of part i have the general form nC_0/n_0 , where C_0 is the sum of the n_0 observations in the control group. Computational symbols (1), (3), (4), and (5) are obtained from the data in part i. The within-cell variation in part ii is obtained from Table 6.7-1.

In comparing the two alternative analyses, it will be noted that the sum of squares for the main effect of factor A is larger in the first analysis than it is in the second analysis (3416 versus 2278). The second type of analysis will always lower this source of variation. Conversely, the alternative analysis will always tend to increase the magnitude of the variation due to interaction. For the numerical data being considered, both analyses lead to essentially the same conclusions.

6.8 Test for Nonadditivity

The assumptions underlying this test were discussed in Sec. 5.20. It should be remembered that only one of several possible sources of nonadditivity is checked in this test—namely, that source which is related to the linear by linear component of an interaction. The numerical details of this test will be considered in this section.

In some cases which arise in practice, a factorial experiment may have only one observation in each cell. If the strictly additive model (no interaction effects) is appropriate, then what are computationally equivalent to

interaction effects may be used as estimates of the experimental error. Computational procedures will be illustrated for the case of a 3×4 factorial experiment having one observation per cell. Suppose that the basic data for this illustration are those given at the left of part i in Table 6.8-1.

Table 6.8-1 Numerical Example

	b_1	b_2	b_3	b_4	Sum	\bar{A}_i	$c_i = \bar{A}_i - \bar{G}$	$\sum_j c_j(AB_{ij}) = d_i$	
(i)	a_1	8	12	16	20	56	14	3	104
	a_2	2	2	14	18	36	9	-2	168
	a_3	5	4	9	22	40	10	-1	166
	Sum	15	18	39	60	132	$11 = \bar{G}$		$\sum_i c_i d_i = -190$
	\bar{B}_j	5	6	13	20	$\sum c_i^2 = 14$			
$c_j = \bar{B}_j - \bar{G}$		-6	-5	2	9	$\sum c_j^2 = 146$			
(1) = 1452		(2) = 1998		(3) = 1508		(4) = 1890			
(ii)	$SS_{\text{nonadd}} = \frac{(\sum c_i d_i)^2}{(\sum c_i^2)(\sum c_j^2)} = \frac{(-190)^2}{(14)(146)} = 17.66$								
(iii)	Source of variation		SS	df	MS	F			
	A		56.00	2		2.57			
	B		438.00	3					
	AB		52.00	6					
	Nonadd		17.66	1	17.66				
	Balance		34.34	5	6.87				

In the column headed \bar{A}_i , the mean of the observations in the corresponding row is entered. In the column headed c_i , an entry has the form $\bar{A}_i - \bar{G}$. For example,

$$c_1 = 14 - 11 = 3,$$

$$c_2 = 9 - 11 = -2.$$

The entries in the row headed c_j have the form $\bar{B}_j - \bar{G}$.

An entry in the column headed d_i is a weighted sum of the entries in the corresponding row, the weights being respective entries in row c_j . For example,

$$d_1 = (-6)(8) + (-5)(12) + (2)(16) + (9)(20) = 104,$$

$$d_2 = (-6)(2) + (-5)(2) + (2)(14) + (9)(18) = 168,$$

$$d_3 = (-6)(5) + (-5)(4) + (2)(9) + (9)(22) = 166.$$

The numerical value of the comparison for nonadditivity is

$$\begin{aligned} \sum_i c_i (\sum_j c_j X_{ij}) &= \sum_i c_i d_i \\ &= (3)(104) + (-2)(168) + (-1)(166) = -190. \end{aligned}$$

The sum of squares corresponding to this comparison is computed in part ii. The numerical values for the computational symbols defined in Table 6.2-1 are also given in part ii.

As a partial check on the computation of SS_{nonadd} , the following relationships must hold:

$$pq(\sum c_i^2)(\sum c_j^2) = (SS_a)(SS_b),$$

$$8(14)(146) = (56)(438),$$

$$24,528 = 24,528.$$

The balance, or residual variation, is given by

$$SS_{\text{bal}} = SS_{ab} - SS_{\text{nonadd}}.$$

Since SS_{nonadd} is a component having a single degree of freedom, SS_{bal} has $(p-1)(q-1)-1$ degrees of freedom. The F ratio in the test for nonadditivity has the form

$$F = \frac{MS_{\text{nonadd}}}{MS_{\text{bal}}} = \frac{17.66}{6.87} = 2.57.$$

If the decision rule leads to accepting the hypothesis of additivity, then the strictly additive model is used in subsequent analyses. On the other hand, if this hypothesis is rejected, the more complete model is used in subsequent analyses. The latter model is generally the more conservative in the sense that higher F ratios are required for significance at a specified level. If the level of significance is set at a numerically high value (say, $\alpha = .25$ rather than $\alpha = .05$ or $.01$), the type 2 error becomes relatively low. In this case low type 2 error implies low probability of using the additive model when in fact it is inappropriate.

If this test is made at the .25 level of significance, the critical value is $F_{.75}(1,5) = 1.69$. Since the observed value of the F statistic exceeds this critical value, the hypothesis of a strictly additive model is rejected at the .25 level of significance. If, however, MS_{ab} is used as the denominator in testing main effects, the resulting test will be biased in the direction of giving too few "significant" results.

There is an alternative method for computing SS_{nonadd} which lends itself more readily to direct generalization to higher-order interaction effects. As a first step in the computation of the comparison desired, one sets up a table in which the entry in cell ij is $c_i c_j$. For the data in Table 6.8-1 the resulting table is as follows:

	b_1	b_2	b_3	b_4	Total
a_1	-18	-15	6	27	0
a_2	12	10	-4	-18	0
a_3	6	5	-2	-9	0
Total	0	0	0	0	

The entry in cell $ab_{11} = (3)(-6) = -18$; the entry $ab_{12} = (3)(-5) = -15$. As a check on the numerical work, each row total must be zero; each column total must also be zero.

The comparison in the test for nonadditivity is a weighted sum of the data in the upper left-hand portion of part i of Table 6.8-1, the weight being the corresponding cell entry in the table given above. Thus,

$$\begin{aligned}\Sigma \Sigma c_i c_j X_{ij} &= (-18)(8) + (-15)(12) + \cdots + (-2)(9) + (-9)(22) \\ &= -190.\end{aligned}$$

As a check on the numerical work,

$$\Sigma \Sigma (c_i c_j)^2 = (\Sigma c_i^2)(\Sigma c_j^2).$$

The sum on the left is given by

$$(-18)^2 + (-15)^2 + \cdots + (-2)^2 + (-9)^2 = 2044.$$

The term on the right is given by

$$(14)(146) = 2044.$$

This latter computational scheme is readily extended to the case of a $p \times q \times r$ factorial experiment. The data in Table 6.8-2 will be used to indicate the numerical details. These data represent a $3 \times 3 \times 3$ factorial experiment in which there is one observation per cell. The usual summary tables for a three-factor factorial experiment are given in part ii. Data for the latter are given in part i. Numerical values of the computational symbols defined in Table 6.5-1 are given in part ii. Since there is only one observation per cell, symbols (2) and (9) are identical, i.e.,

$$\Sigma X_{ijk}^2 = \Sigma (ABC)_{ijk}^2.$$

(The letter c is used for two different concepts, but the context should make clear what is meant. In one context $c_1, c_2,$ and c_3 represent the levels of factor C . In a second context $c_i, c_j,$ and c_k represent deviations from the grand mean.)

Means and the deviations of the means from the grand mean are computed in part iv. In this context $c_k = \bar{C}_k - \bar{G}$. Similarly, $c_i = \bar{A}_i - \bar{G}$. The entry in cell abc_{ijk} in part v has the general form $c_i c_j c_k$. Thus the entry in cell abc_{111} is $(-3.4)(-1.8)(-5.1) = -31.2$. The entry in cell abc_{123} is $(-3.4)(1.2)(3.2) = -13.1$. As a check on the computational work,

$$\sum_i c_i c_j c_k = \sum_j c_i c_j c_k = \sum_k c_i c_j c_k = 0.$$

That is, the sum of any column in part v must be zero; also the sum over b_j within any fixed level of ac_{ik} must be zero. For example, for level ac_{11}

$$-31.2 + 20.8 + 10.4 = 0.0.$$

For level ac_{12} ,

$$11.6 + (-7.7) + (-3.9) = 0.0.$$

Table 6.8-2 Numerical Example

(i)		c_1			c_2			c_3			Total
		b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3	
	a_1	3	6	9	6	9	12	9	12	15	81
	a_2	6	9	12	12	18	21	15	21	21	135
a_3	9	9	3	18	21	12	18	18	12	120	
		18	24	24	36	48	45	42	51	48	336

AB summary table

	b_1	b_2	b_3	Total
a_1	18	27	36	81
a_2	33	48	54	135
a_3	45	48	27	120
	96	123	117	336

AC summary table

	c_1	c_2	c_3	Total
a_1	18	27	36	81
a_2	27	51	57	135
a_3	21	51	48	120
	66	129	141	336

BC summary table

	c_1	c_2	c_3	Total
b_1	18	36	42	96
b_2	24	48	51	123
b_3	24	45	48	117
	66	129	141	336

	(1) = 4181.33	(4) = 4226.00	(7) = 4758.00
(iii)	(2) = 4986.00	(5) = 4542.00	(8) = 4590.00
	(3) = 4354.00	(6) = 4572.00	(9) = 4986.00

	$\bar{G} = 12.4$	\bar{A}_i	\bar{B}_j	\bar{C}_k	c_i	c_j	c_k	$\Sigma c_i^2 = 18.62$
(iv)	1	9	10.7	7.3	-3.4	-1.8	-5.1	$\Sigma c_j^2 = 5.04$
	2	15	13.7	14.3	2.5	1.2	1.9	$\Sigma c_k^2 = 39.86$
	3	13.3	13	15.7	0.9	0.6	3.2	

[illegible]

The comparison associated with nonadditivity is a weighted sum of the entries in the cell of part i , the weights being the corresponding entries in part v . Thus the comparison used in the test for nonadditivity is

$$(-31.2)(3) + (20.8)(6) + (10.4)(9) + \cdots + (1.7)(12) = 60.60.$$

The sum of squares for this comparison has the form

$$SS_{\text{nonadd}} = \frac{(60.60)^2}{\sum (c_i c_j c_k)^2}.$$

The divisor is given by

$$\sum (c_i c_j c_k)^2 = (-31.2)^2 + (20.8)^2 + \cdots + (1.7)^2 = 3742.$$

Within rounding error, the following relation must hold:

$$\sum (c_i c_j c_k)^2 = (\sum c_i^2)(\sum c_j^2)(\sum c_k^2).$$

The right-hand side of this last expression is

$$(18.62)(5.04)(39.86) = 3740.7.$$

The analysis of variance is summarized in Table 6.8-3.

Table 6.8-3 Summary of Analysis of Variance

Source of variation	SS	df	MS	F
<i>A</i>	172.67	2		
<i>B</i>	44.67	2		
<i>C</i>	360.67	2		
<i>AB</i>	173.33	4		
<i>AC</i>	43.33	4		
<i>BC</i>	3.33	4		
<i>ABC</i>	6.67	8		
Nonadd 0.98		1	0.98	1.21
Balance 5.69		7	0.81	

The test for nonadditivity is given by

$$F = \frac{0.98}{0.81} = 1.21.$$

The critical value of this statistic for a .25-level test is $F_{.75}(1,7) = 1.57$. Since the observed value of the F statistic does not exceed this critical value, there is no reason to reject the hypothesis of additivity. The evidence from this test supports the hypothesis that the components of the three-factor interaction are homogeneous. If the assumptions underlying the test are met, the component for nonadditivity would tend to be large relative to the other components, provided that the three-factor interaction estimated a source of variation different from experimental error. In this case, the three-factor interaction may be considered as an estimate of experimental

error (granting the validity of the assumptions). Hence the three-factor interaction term may be dropped from the model. In the latter case, MS_{abe} provides an estimate of σ_e^2 .

6.9 Computation of Trend Components

Computational procedure for trends for the case of a single-factor experiment were discussed in Sec. 3.7. These procedures generalize to factorial experiments. Principles underlying this generalization were discussed in Sec. 5.18; the actual computation of trend components in a factorial experiment are considered in this section. Computational procedures will be illustrated for the case of a 3×4 factorial experiment having five observations in each cell.

It will be assumed that (1) both factors are fixed, (2) the levels of both factors represent steps along an underlying quantitative scale, and (3) the respective levels represent equally spaced steps along the respective scales. The latter assumption, which is not essential to the development, permits a simplification of the numerical work, since coefficients of orthogonal polynomials may be used to obtain desired sums of squares. For this case, the coefficients for the levels of factor A are as follows (see Table B.10):

Linear: c'_i	-1	0	1
Quadratic: c''_i	1	-2	1

The coefficients for the levels of factor B are as follows:

Linear: c'_j	-3	-1	1	3
Quadratic: c''_j	1	-1	-1	1
Cubic: c'''_j	-1	3	-3	1

The data given in part i of Table 6.9-1 will be used as a numerical example. [Assume that each entry in the AB summary table is the sum of five observations; data for the computation of symbol (2) are not given.] The analysis of variance is summarized in part iii. This particular analysis does not necessarily give the experimenter all the information he seeks. There are many other ways in which the over-all variation may be analyzed.

In spite of the significant interaction, it may be of interest to study the trend components of the main effects. For illustrative purposes, the trend components of the B main effect will be obtained. The comparison associated with the linear component is a weighted sum of B_j totals, the weights being the linear coefficients. For data in part i of Table 6.9-1, the linear comparison is

$$C_{11n} = (-3)(19) + (-1)(21) + (1)(28) + (3)(37) = 61.$$

Table 6.9-1 Numerical Example

AB summary table ($n = 5$):

		b_1	b_2	b_3	b_4	Total
(i)	a_1	3	5	9	14	31
	a_2	7	11	15	20	53
	a_3	9	5	4	3	21
		19	21	28	37	105

(ii)	(1) = 183.75	(3) = 210.55	(5) = 247.40
	(2) = 280.00	(4) = 197.00	

	Source of variation	SS	df	MS	F
(iii)	A	26.80	2	13.40	19.71
	B	13.25	3	4.42	6.50
	AB	23.60	6	3.93	5.78
	Within cell	32.40	48	0.68	

The linear component of the variation due to the main effects of factor B is

$$SS_{b(\text{lin})} = \frac{C_{\text{lin}}^2}{np\Sigma(c'_j)^2} = \frac{(61)^2}{5(3)(20)} = 12.40.$$

Computational formulas for the quadratic and cubic components of the B main effect are summarized in Table 6.9-2. It is noted that the linear component accounts for $12.40/13.25$, or 94 per cent, of the variation due to the main effect. This means that on the average over levels of factor A the criterion measure predominantly is a linear function of levels of factor A .

Table 6.9-2 Trends of B Main Effects

	c'_j	-3	-1	1	3	$\Sigma(c'_j)^2 = 20$
(i)	c''_j	1	-1	-1	1	$\Sigma(c''_j)^2 = 4$
	c'''_j	-1	3	-3	1	$\Sigma(c'''_j)^2 = 20$

	B_j :	19	21	28	37
--	---------	----	----	----	----

(ii)	$SS_{b(\text{lin})}$	$= \frac{(\Sigma c'_j B_j)^2}{np\Sigma(c'_j)^2} = \frac{(61)^2}{5(3)(20)} = 12.40$
	$SS_{b(\text{quad})}$	$= \frac{(\Sigma c''_j B_j)^2}{np\Sigma(c''_j)^2} = \frac{(7)^2}{5(3)(4)} = 0.82$
	$SS_{b(\text{cubic})}$	$= \frac{(\Sigma c'''_j B_j)^2}{np\Sigma(c'''_j)^2} = \frac{(-3)^2}{5(3)(20)} = 0.03$
	SS_b	$= 13.25$

A test on whether a trend component differs significantly from zero uses the statistic

$$F = \frac{MS_{\text{trend}}}{MS_{w, \text{cell}}}.$$

For example, a test on the quadratic trend is given by

$$F = \frac{MS_{b(\text{quad})}}{MS_{w, \text{cell}}} = \frac{0.82}{0.68} = 1.21.$$

The critical value for a .05-level test is $F_{.95}(1, 48) = 4.04$. Hence the experimental data indicate that the hypothesis of no quadratic trend in the main effect of factor B is tenable.

Table 6.9-3 Trends within AB Interaction

		Linear \times Linear				
		b_1	b_2	b_3	b_4	
(i)	a_1	3	1	-1	-3	$SS_{\text{lin} \times \text{lin}} = \frac{[\sum d_{ij}(AB_{ij})]^2}{n(\sum d_{ij}^2)}$ $= \frac{(-56)^2}{5(40)} = 15.68$
	a_2	0	0	0	0	
	a_3	-3	-1	1	3	
		Quadratic \times Linear				
		b_1	b_2	b_3	b_4	
(ii)	a_1	-3	-1	1	3	$SS_{\text{quad} \times \text{lin}} = \frac{[\sum d_{ij}(AB_{ij})]^2}{n(\sum d_{ij}^2)}$ $= \frac{(-66)^2}{5(120)} = 7.26$
	a_2	6	2	-2	-6	
	a_3	-3	1	-1	3	
		Linear \times Quadratic				
		b_1	b_2	b_3	b_4	
(iii)	a_1	-1	1	1	-1	$SS_{\text{lin} \times \text{quad}} = \frac{[\sum d_{ij}(AB_{ij})]^2}{n(\sum d_{ij}^2)}$ $= \frac{(0)^2}{5(8)} = 0$
	a_2	0	0	0	0	
	a_3	1	-1	-1	1	

A significant interaction implies that the response surface for different levels of factor B (or A) is not homogeneous, i.e., that profiles are not parallel. The linear \times linear, linear \times quadratic, etc., components of interaction indicate the fit of variously shaped surfaces, i.e., different patterns of profiles. Computational procedures are summarized in Table 6.9-3. The weights for the linear \times linear comparison are given in part i. An entry in cell ab_{ij} of this table has the form

$$d_{ij} = c'_i c'_j.$$

For example,

$$\begin{aligned} d_{11} &= (-1)(-3) = 3, & d_{12} &= (-1)(-1) = 1, \\ d_{21} &= (0)(-3) = 0, & d_{22} &= (0)(-3) = 0. \end{aligned}$$

The weights for the quadratic \times linear comparison are given in part ii. An entry in this latter table is given by

$$d_{ij} = c_i' c_j'.$$

(Although the symbol d_{ij} is used for the typical entry in different tables, the context will make it clear which table is meant. A more explicit, but more cumbersome, notation for the latter d_{ij} would be $d_{ij'}$.) The entries in part ii are obtained as follows:

$$\begin{aligned} d_{11} &= (1)(3) = 3, & d_{12} &= (1)(-1) = -1, \\ d_{21} &= (-2)(3) = -6, & d_{22} &= (-2)(-1) = 2, \\ d_{31} &= (1)(3) = 3, & d_{32} &= (1)(-1) = -1. \end{aligned}$$

Computational formulas for some of the trend components of the AB interaction are summarized in Table 6.9-3. With suitable definition of d_{ij} , the other components have the same general form. In each case d_{ij} refers to an entry in a different table of weights. Of the total variation due to AB , the linear \times linear component accounts for 15.68/23.60, or 66 per cent. The sum of the linear \times linear component and the quadratic \times linear components accounts for

$$\frac{15.68 + 7.26}{23.60} = .97,$$

or 97 per cent. Tests on trend components of the interaction have the following general form:

$$F = \frac{MS_{\text{trend}}}{MS_{w, \text{cell}}}.$$

Table 6.9-4 Difference in Trends for Simple Effects of Factor B

	b_1	b_2	b_3	b_4	Total	d_i'	d_i''	d_i'''
a_1	3	5	9	14	31	37	3	-1
a_2	7	11	15	20	53	43	1	1
a_3	9	5	4	3	21	-19	3	-3
(i)	19	21	28	37	105	61	7	-3
c_j'	-3	-1	1	3	$\Sigma(c_j')^2 = 20$			
c_j''	1	-1	-1	1	$\Sigma(c_j'')^2 = 4$			
c_j'''	-1	3	-3	1	$\Sigma(c_j''')^2 = 20$			
<hr/>								
(ii)								
$SS_{\text{diff in lin trend}} = \frac{\Sigma(d_i')^2}{n\Sigma(c_j')^2} - \frac{(\Sigma d_i')^2}{np\Sigma(c_j')^2} = 35.79 - 12.40 = 23.39$								
$SS_{\text{diff in quad trend}} = \frac{\Sigma(d_i'')^2}{n\Sigma(c_j'')^2} - \frac{(\Sigma d_i'')^2}{np\Sigma(c_j'')^2} = .95 - .82 = .13$								
$SS_{\text{diff in cubic trend}} = \frac{\Sigma(d_i''')^2}{n\Sigma(c_j''')^2} - \frac{(\Sigma d_i''')^2}{np\Sigma(c_j''')^2} = .11 - .03 = .08$								
						$SS_{\text{ab}} = 23.60$		

It is sometimes of interest to study differences in trends for the simple effects of one factor at different levels of a second factor. For illustrative purposes, differences between trends for the simple effects of factor B at different levels of factor A will be obtained. (The degrees of freedom for such differences in trend are $p - 1 = 2$ for each trend.) Computational formulas for these sources of variation are given in Table 6.9-4.

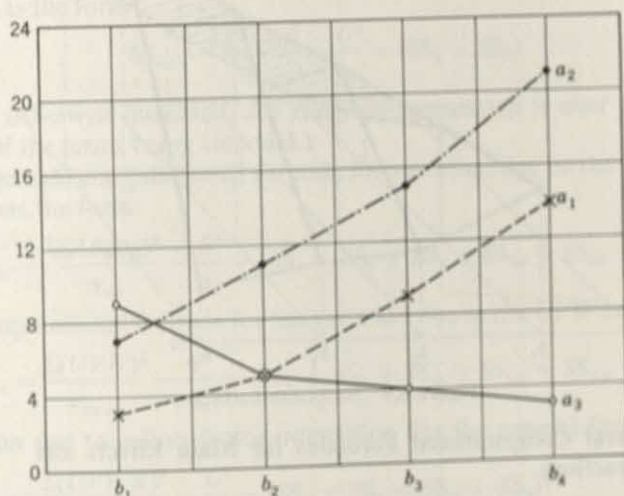


Figure 6.8 Profiles of factor B at levels of factor A .

In this table the symbol d'_i is defined as follows:

$$d'_i = \sum_j c'_j(AB_{ij}).$$

For example,

$$d'_1 = (-3)(3) + (-1)(5) + (1)(9) + (3)(14) = 37.$$

The symbols d''_i and d'''_i are defined as follows:

$$d''_i = \sum_j c''_j(AB_{ij}),$$

$$d'''_i = \sum_j c'''_j(AB_{ij}).$$

The variation due to differences in linear trends in simple effects of factor B explains 23.39/23.60, or 99 per cent, of the total variation of the AB interaction. This means that 99 per cent of the AB interaction arises from differences between the linear trends in the profiles of factor B at the different levels of factor A . These profiles are shown in Fig. 6.8. If these profiles were plotted in a three-dimensional space, the response surface represented by the AB summary table would be obtained. This response surface is shown in Fig. 6.9. From this surface it is seen that profiles for

factor A at fixed levels of factor B are predominantly quadratic in form. However, the profiles for factor B at fixed levels of factor A tend to be linear in form. Hence the shape of this surface is predominantly quadratic \times linear.

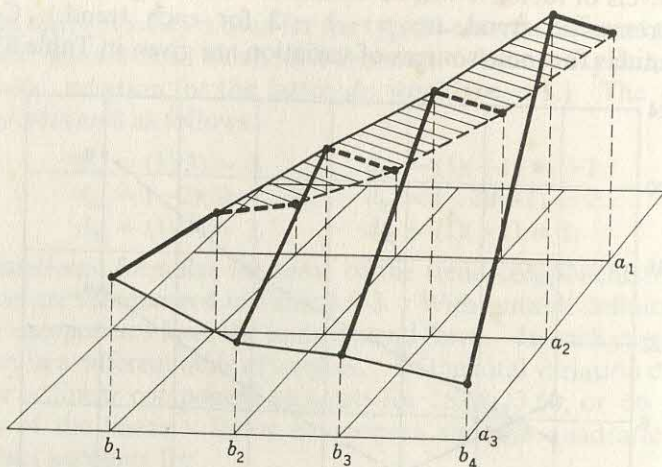


Figure 6.9 Response surface.

6.10 General Computational Formulas for Main Effects and Interactions

The following notation will be used in this section.

A_i = sum of all observations at level a_i .

n_a = number of observations summed to obtain A_i ; n_a is assumed constant for all levels of factor A .

AB_{ij} = sum of all observations at level ab_{ij} .

n_{ab} = number of observations summed to obtain AB_{ij} ; n_{ab} is assumed constant for all ab_{ij} 's.

ABC_{ijk} = sum of all observations at level abc_{ijk} .

n_{abc} = number of observations summed to obtain ABC_{ijk} ; n_{abc} is assumed constant for all abc_{ijk} 's.

In terms of this notation, a general formula for the main effect due to factor A is

$$SS_a = \frac{\sum A_i^2}{n_a} - \frac{G^2}{n_g},$$

where G is the grand total of all observations and n_g is the number of observations summed to obtain G . (In this context an observation is a measurement on the experimental unit.) A general formula for the main effect due to factor B has the form

$$SS_b = \frac{\sum B_j^2}{n_b} - \frac{G^2}{n_g}.$$

The general computational formula for the variation due to the AB interaction is

$$SS_{ab} = \frac{\Sigma(AB_{ij})^2}{n_{ab}} - \frac{G^2}{n_g} - (SS_a + SS_b).$$

The general computational formula for the variation due to the AC interaction has the form

$$SS_{ac} = \frac{\Sigma(AC_{ij})^2}{n_{ac}} - \frac{G^2}{n_g} - (SS_a + SS_c).$$

(Unless otherwise indicated, the range of summation is over all possible values of the terms being summed.)

The general computational formula for variation due to the ABC interaction has the form

$$SS_{abc} = \frac{\Sigma(ABC_{ijk})^2}{n_{abc}} - \frac{G^2}{n_g} - (SS_a + SS_b + SS_c + SS_{ab} + SS_{ac} + SS_{bc}).$$

The computational formula for the variation due to the UVW interaction is

$$SS_{uvw} = \frac{\Sigma(UVW)^2}{n_{uvw}} - \frac{G^2}{n_g} - (SS_u + SS_v + SS_w + SS_{uv} + SS_{uw} + SS_{vw}).$$

Variation due to a four-factor interaction has the general formula

$$\begin{aligned} SS_{uvwx} = \frac{\Sigma(UVWX)^2}{n_{uvwx}} - \frac{G^2}{n_g} - (SS_u + SS_v + SS_w + SS_x) \\ - (SS_{uv} + SS_{uw} + SS_{ux} + SS_{vw} + SS_{vx} + SS_{wx}) \\ - (SS_{uvw} + SS_{uvx} + SS_{uwx} + SS_{vwx}). \end{aligned}$$

In a factorial experiment having t factors there are

$$\binom{t}{2} \text{ two-factor interactions,}$$

$$\binom{t}{3} \text{ three-factor interactions,}$$

.....,

$$\binom{t}{m} \text{ } m\text{-factor interactions.}$$

For example, in a four-factor experiment, the number of possible two-factor interactions is

$$\binom{4}{2} = \frac{4(3)}{1(2)} = 6.$$

In a four-factor experiment the number of possible three-factor interactions is

$$\binom{4}{3} = \frac{4(3)(2)}{1(2)(3)} = 4.$$

Alternatively, given a t -factor interaction, the formulas that have been given above indicate the number of different two-factor, three-factor, etc., interactions that may be formed from the t letters in the t -factor interaction.

The general formula for variation due to a five-factor interaction is

$$SS_{uvwx y} = \frac{\Sigma(UVWXY)^2}{n_{uvwx y}} - \frac{G^2}{n_g} - [(i) + (ii) + (iii) + (iv)]$$

where (i) = $\Sigma SS_{\text{main effect}}$ —there are five terms in this sum;

(ii) = $\Sigma SS_{\text{two-factor int}}$ —there are $[5(4)]/[1(2)] = 10$ terms in this sum;

(iii) = $\Sigma SS_{\text{three-factor int}}$ —there are $[5(4)(3)]/[1(2)(3)] = 10$ terms in this sum;

(iv) = $\Sigma SS_{\text{four-factor int}}$ —there are $[5(4)(3)(2)]/[1(2)(3)(4)] = 5$ terms in this sum.

The general formulas given above assume that no factor is nested under any other. In case factor C is nested under factor B , the general formula for the main effect due to C within B is

$$SS_{C \text{ w. } B} = \frac{\Sigma(BC)^2}{n_{bc}} - \frac{\Sigma B^2}{n_b} \\ = SS_c + SS_{bc},$$

where SS_c is the variation due to the main effect of factor C if the nesting is disregarded and SS_{bc} has an analogous definition. The general formula for the $A \times (C \text{ within } B)$ variation is

$$SS_{A(c \text{ w. } b)} = \frac{\Sigma(ABC)^2}{n_{abc}} - \frac{\Sigma(AB)^2}{n_{ab}} - \frac{\Sigma(BC)^2}{n_{bc}} + \frac{\Sigma B^2}{n_b} \\ = SS_{ac} + SS_{abc},$$

where SS_{abc} is the three-factor variation computed with the nesting disregarded. In general,

$$SS_{uv(x \text{ w. } w)} = SS_{uvx} + SS_{uvzw},$$

where the terms on the right are computed as if there were no nesting.

For the case in which factor C is nested under factor B and factor D is nested under factor C ,

$$SS_{d(w, b \text{ and } c)} = \frac{\Sigma(BCD)^2}{n_{bcd}} - \frac{\Sigma(BC)^2}{n_{bc}} \\ = SS_d + SS_{bd} + SS_{cd} + SS_{bcd}.$$

The terms in this last line are computed as if there were no nesting. Also,

$$SS_{ad(w, b \text{ and } c)} = \frac{\Sigma(ABCD)^2}{n_{abcd}} - \frac{\Sigma(ABC)^2}{n_{abc}} - \frac{\Sigma(BCD)^2}{n_{bcd}} + \frac{\Sigma(BC)^2}{n_{bc}} \\ = SS_{ad} + SS_{abd} + SS_{acd} + SS_{abcd}.$$

The degrees of freedom for the variation on the left are the sum of the degrees of freedom of the variations on the right. The degrees of freedom may be checked as follows:

$$\begin{aligned} df &= pqs - pqr - qrs + qr \\ &= pqr(s - 1) - qr(s - 1) \\ &= qr(s - 1)(p - 1). \end{aligned}$$

The terms in the first line of the above expression are obtained as follows.

pqs = number of observations in an $ABCD$,

pqr = number of observations in an ABC ,

qrs = number of observations in a BCD ,

pq = number of observations in a BC .

6.11 Missing Data

For a factorial experiment in which the cell frequencies are not equal but all n_{ij} 's are approximately equal, the analysis by the method of unweighted means presents no particular problems, provided that all cells contain at least one observation. In the unweighted-means analysis, the mean of each cell is estimated by the observations actually made within that cell. In cases where the number of observations within a particular cell is small relative to the number of observations in other cells in the same row and column or adjacent rows and adjacent columns, then information provided by these other cells may be used in estimating the mean of a specified cell. In an unweighted-means analysis, such information is not utilized in the estimation process.

In an experiment having no observed data in given cells, estimates of such cell means may have to be obtained from other cells in the experiment. If the form of the response surface were known, estimates of the missing entries could be obtained by using a multiple regression equation. This solution is generally not practical. One method, having somewhat limited utility, is to estimate the missing mean in cell ij by the following formula,

$$\overline{AB}_{ij}' = \bar{A}_i' + \bar{B}_j' - \bar{G}',$$

where \bar{A}_i' is the mean of the observations actually made under level a_i ; similar definitions hold for \bar{B}_j' and \bar{G}' . This method of estimating missing data assumes no interaction present. Because this method of estimation does not take into account trends in the row and column of the missing entry, its utility is limited. For the case of a three-factor experiment, the method that has just been described takes the form

$$\overline{ABC}_{ijk}' = \bar{A}_i' + \bar{B}_j' + \bar{C}_k' - 2\bar{G}'.$$

The presence of interaction effects is a stumbling block in all methods for estimating missing cell entries. A review of methods that have been proposed and an extensive bibliography on the topic of missing data are given

in Federer (1955, pp. 124–127, 133–134). Caution and judgment are called for in the use of any method for estimating missing data. Each of the proposed methods has a set of assumptions that must be considered with care.

Consider the numerical data given in Table 6.11-1. These data represent the available observations in a 5×4 factorial experiment. Assume the levels of the factors represent steps along a quantitative scale. There is a single measurement in 18 of the 20 cells. Entries in cells ab_{21} and ab_{34} are missing. The unknown values of these cell entries are designated u_{21} and u_{34} . [The data in this table appear in Bennett and Franklin (1954, p. 383).] A method of estimating missing observations, which minimizes the

Table 6.11-1 Numerical Example

	b_1	b_2	b_3	b_4	Total
a_1	28	20	11	10	69
a_2	u_{21}	16	15	8	$39 + u_{21}$
a_3	29	13	16	u_{34}	$58 + u_{34}$
a_4	27	10	18	11	66
a_5	28	11	15	10	64
	112	70	75	39	$296 + u_{21} + u_{34}$
	+			+	
	u_{21}			u_{34}	

interaction effect in the analysis resulting when such estimates are used in place of the missing data, is described by Bennett and Franklin (1954, pp. 382–383). Using this method, Bennett and Franklin arrive at the following estimates:

$$u_{21} = 28.6, \quad u_{34} = 10.2.$$

If these two estimates of the missing entries are inserted in their proper places and the usual analysis of variance computed as if all data were observed, the variation due to AB will be a minimum. In essence, this principle for estimating the missing entries keeps the profiles of the simple effects as parallel as possible.

A simplified version of this latter principle utilizes information from only those cells adjacent to the cell in which data are missing. For example,

$$\frac{u_{21}}{16} = \frac{\left(\frac{28}{20}\right) + \left(\frac{29}{13}\right)}{2}.$$

Solving for u_{21} gives

$$u_{21} = 29.0.$$

This method of estimation assumes that the slope of the profiles for levels b_1 and b_2 at level a_2 is the mean of corresponding slopes at levels a_1 and a_3 . The latter two slopes can be obtained from the observed data. Analogously,

$$\frac{u_{34}}{16} = \frac{\left(\frac{8}{15}\right) + \left(\frac{11}{18}\right)}{2},$$

from which

$$u_{34} = 9.2.$$

To use the first method discussed in this section for estimating missing entries, one proceeds as follows:

$$\begin{aligned}\bar{A}'_2 &= \frac{39}{3} = 13.0, & \bar{A}'_3 &= 19.3; \\ \bar{B}'_1 &= \frac{112}{4} = 28.0, & \bar{B}'_4 &= 9.8; \\ \bar{G}' &= \frac{296}{18} = 16.4.\end{aligned}$$

In each case the denominator is the number of observations in the corresponding total. The estimates of the unknown entries are

$$\begin{aligned}u_{21} &= 13.0 + 28.0 - 16.4 = 24.6, \\ u_{34} &= 19.3 + 9.8 - 16.4 = 12.6.\end{aligned}$$

If the analysis of variance were to be carried out with estimated values substituted for the missing values, the degrees of freedom for the resulting two-factor interaction would be

$$df_{ab} = (p - 1)(q - 1) - (\text{number of missing values}).$$

For the numerical example in Table 6.11-1,

$$df_{ab} = 4(3) - 2 = 10.$$

By way of summary, there are mathematically elegant methods for estimating missing cell entries. None of these methods is satisfactory unless the experimenter has information about the nature of the response surface being studied. There is, however, no real substitute for experimental data. In the example given in Table 6.11-1, the experimentally determined values of the missing entries were

$$u_{21} = 23, \quad u_{34} = 24.$$

None of the methods considered yielded values relatively close to the observed value for u_{34} .

6.12 Special Computational Procedures When All Factors Have Two Levels

When each factor in a factorial experiment has two levels, the computational formulas for main effects and interactions may be simplified. For the case of a 2×2 factorial experiment having n observations per cell,

$$\begin{aligned}SS_a &= \frac{(A_1 - A_2)^2}{4n}, \\ SS_b &= \frac{(B_1 - B_2)^2}{4n}, \\ SS_{ab} &= \frac{[(AB_{11} + AB_{22}) - (AB_{12} + AB_{21})]^2}{4n}.\end{aligned}$$

In the expression for the interaction, note that the sum of the subscripts of each term in the first set of parentheses is an even number, whereas the sum of the subscripts for each term in the second set of parentheses is an odd number.

For the case of a $2 \times 2 \times 2$ factorial experiment having n observations per cell,

$$SS_a = \frac{(A_1 - A_2)^2}{8n},$$

$$SS_b = \frac{(B_1 - B_2)^2}{8n},$$

$$SS_c = \frac{(C_1 - C_2)^2}{8n},$$

$$SS_{ab} = \frac{[(AB_{11} + AB_{22}) - (AB_{12} + AB_{21})]^2}{8n},$$

$$SS_{ac} = \frac{[(AC_{11} + AC_{22}) - (AC_{12} + AC_{21})]^2}{8n},$$

$$SS_{bc} = \frac{[(BC_{11} + BC_{22}) - (BC_{12} + BC_{21})]^2}{8n},$$

$$SS_{abc} = \frac{[(ABC_{111} + ABC_{122} + ABC_{212} + ABC_{221}) - (ABC_{112} + ABC_{121} + ABC_{211} + ABC_{222})]^2}{8n}.$$

Again note that in expressions for interactions the sum of the subscripts for each term within a pair of parentheses is an odd number in one case and an even number in the other case.

The computations for a three-factor experiment may be made without obtaining two-factor summary tables by following the scheme given in Table 6.12-1.

The ABC summary table is given in part i. The symbol s'_1 in part i' represents the sum of the entries in the first column of part i; s'_2 , s'_3 , and s'_4 represent sums of entries in the respective columns of part i. These sums should be arranged as indicated in part i'. The symbol d'_1 represents the difference between the two entries in the first column in part i, that is, $d'_1 = ABC_{111} - ABC_{211}$. The symbols d'_2 , d'_3 , and d'_4 represent corresponding differences between entries in corresponding column in part i. For example, $d'_4 = ABC_{122} - ABC_{222}$.

The entries in part i'' are obtained from the entries in part i' in the same general manner as corresponding entries were obtained from part i. That is,

$$s''_1 = s'_1 + s'_2, \quad s''_3 = d'_1 + d'_2,$$

$$s''_2 = s'_3 + s'_4, \quad s''_4 = d'_3 + d'_4.$$

Also,

$$d''_1 = s'_1 - s'_2, \quad d''_3 = d'_1 - d'_2,$$

$$d''_2 = s'_3 - s'_4, \quad d''_4 = d'_3 - d'_4.$$

The entries in part i''' are obtained from the entries in part i'' by means of the same general pattern.

$$s_1''' = s_1'' + s_2'', \quad s_3''' = d_1'' + d_2'',$$

$$s_2''' = s_3'' + s_4'', \quad s_4''' = d_3'' + d_4''.$$

Also,

$$d_1''' = s_1'' - s_2'', \quad d_3''' = d_1'' - d_2'',$$

$$d_2''' = s_3'' - s_4'', \quad d_4''' = d_3'' - d_4''.$$

Table 6.12-1 Computational Scheme for $2 \times 2 \times 2$ Factorial Experiment

	c_1		c_2	
	b_1	b_2	b_1	b_2
(i)	a_1	ABC_{111}	ABC_{121}	ABC_{112}
	a_2	ABC_{211}	ABC_{221}	ABC_{212}
		(1)	(2)	(3)
				(4)
(i')		s_1'	s_3'	d_1'
		s_2'	s_4'	d_2'
(i'')		s_1''	s_3''	d_1''
		s_2''	s_4''	d_2''
(i''')		s_1'''	s_3'''	d_1'''
		s_2'''	s_4'''	d_2'''
(ii)		$SS_a = (s_2''')^2/8n$	$SS_c = (d_1''')^2/8n$	
		$SS_b = (s_3''')^2/8n$	$SS_{ac} = (d_2''')^2/8n$	
		$SS_{ab} = (s_4''')^2/8n$	$SS_{bc} = (d_3''')^2/8n$	
			$SS_{abc} = (d_4''')^2/8n$	

Computational formulas for the sums of squares are given in part ii. These computational formulas are identical to those given earlier in this section. A numerical example of this computational scheme is given in Table 6.12-2. As a check on the computational work,

$$\begin{aligned}
 SS_{\text{between cells}} &= \frac{\Sigma(ABC)^2}{n} - \frac{G^2}{8n} \\
 &= \Sigma(\text{main effects}) + \Sigma(\text{two-factor int}) \\
 &\quad + \Sigma(\text{three-factor int}).
 \end{aligned}$$

For the data in Table 6.12-2,

$$SS_{\text{between cells}} = 625.00 - 525.62 = 99.38.$$

Table 6.12-2 Numerical Example

	c_1		c_2		$(n = 5)$
	b_1	b_2	b_1	b_2	
(i)					
	a_1	10	15	20	10
	a_2	20	30	30	10
		(1)	(2)	(3)	(4)
<hr/>					
(i')		30	50	-10	-10
		45	20	-15	0
<hr/>					
(i'')		75	-25	-15	5
		70	-10	30	-10
<hr/>					
(i''')		145	15	5	-45
		-35	-5	-15	15
<hr/>					
(ii)	$SS_a = (-35)^2/40 = 30.62$	$SS_c = (5)^2/40 = .62$			
	$SS_b = (15)^2/40 = 5.62$	$SS_{ac} = (-15)^2/40 = 5.62$			
	$SS_{ab} = (-5)^2/40 = .62$	$SS_{bc} = (-45)^2/40 = 50.62$			
		$SS_{abc} = (15)^2/40 = 5.62$			

Within rounding error, this is the numerical value of the sum of the variations computed in part ii.

The computational scheme that has just been illustrated may be generalized to any 2^k factorial experiment, where k is the number of factors. The generalization to a $2^4 = 2 \times 2 \times 2 \times 2$ factorial experiment is illustrated in Table 6.12-3. In part i the $ABCD$ summary table is represented schematically. In part i' each s' represents the sum of the entries in the corresponding columns of part i; each d' represents the difference between the elements in the corresponding columns of part i. Entries in part i'' are obtained from the entries in the columns of part i' in an analogous manner. That is,

$$s''_1 = s'_1 + s'_2, \quad d''_1 = s'_1 - s'_2;$$

$$s''_2 = s'_3 + s'_4, \quad d''_2 = s'_3 - s'_4;$$

$$\dots \dots \dots, \quad \dots \dots \dots;$$

$$s''_8 = d'_7 + d'_8, \quad d''_8 = d'_7 - d'_8.$$

This procedure continues until part i^k is completed, in this case until part i^{iv} is completed. Computational formulas for the variation due to the sources indicated in part ii may be obtained from the corresponding entry in part i^{iv} . For example,

$$SS_c = \frac{(s_5^{iv})^2}{16n}, \quad SS_{abd} = \frac{(d_4^{iv})^2}{16n}.$$

Table 6.12-3 Computational Scheme for 2^4 Factorial Experiment

(i)	a_1 a_2	d_1				d_2			
		c_1		c_2		c_1		c_2	
		b_1	b_2	b_1	b_2	b_1	b_2	b_1	b_2
		(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(i')		s'_1	s'_3	s'_5	s'_7	d'_1	d'_3	d'_5	d'_7
		s'_2	s'_4	s'_6	s'_8	d'_2	d'_4	d'_6	d'_8
(i'')		s''_1	s''_3	s''_5	s''_7	d''_1	d''_3	d''_5	d''_7
		s''_2	s''_4	s''_6	s''_8	d''_2	d''_4	d''_6	d''_8
(i''')		s'''_1	s'''_3	s'''_5	s'''_7	d'''_1	d'''_3	d'''_5	d'''_7
		s'''_2	s'''_4	s'''_6	s'''_8	d'''_2	d'''_4	d'''_6	d'''_8
(i'iv)		s^{iv}_1	s^{iv}_3	s^{iv}_5	s^{iv}_7	d^{iv}_1	d^{iv}_3	d^{iv}_5	d^{iv}_7
		s^{iv}_2	s^{iv}_4	s^{iv}_6	s^{iv}_8	d^{iv}_2	d^{iv}_4	d^{iv}_6	d^{iv}_8
(ii)		—	B	C	BC	D	BD	CD	BCD
		A	AB	AC	ABC	AD	ABD	ACD	$ABCD$

The general formula is

$$\frac{(s^k)^2}{2^k n} \quad \text{or} \quad \frac{(d^k)^2}{2^k n}$$

This computational procedure is particularly useful when k is 4 or larger.

6.13 Illustrative Applications

In a study reported by Wulff and Stolurow (1957) the experimental plan was a 2×2 factorial having 10 subjects per cell. Factor A was essentially a classification factor; its levels indicated the aptitude of the subjects as measured by a test of mechanical aptitude. The levels of factor B were the methods of instruction used in a paired-associates learning task. The criterion was the number of correct responses in a block of trials. The analysis of variance had the following form:

Source	df	MS	F
A (aptitude)	1	2175.62	17.96
B (method)	1	931.22	7.68
AB	1	9.02	
Within cell	36	121.10	

Factors A and B were both considered to be fixed; hence $MS_{w, cell}$ was the denominator for all F ratios. The factor of primary interest was B ; the critical value for a .05-level test on this factor is $F_{.95}(1,36) = 4.11$. The experimental data indicated that there is a statistically significant difference between the two methods of instruction with respect to the mean number of correct responses in the specified blocks of trials.

Gordon (1959) reports a 4×4 factorial experiment in which both factors A and B were considered to be random. The levels of factor A represented the kind of task on which subjects were given pretraining; the levels of factor B represented the kind of task to which the subjects were transferred. The design may be represented schematically as follows:

	b_1	b_2	b_3	b_4
a_1	G_{11}	G_{12}	G_{13}	G_{14}
a_2	G_{21}	G_{22}	G_{23}	G_{24}
a_3	G_{31}	G_{32}	G_{33}	G_{34}
a_4	G_{41}	G_{42}	G_{43}	G_{44}

The symbol G_{ij} represents the group of subjects given pretraining under task i and transferred to task j . The criterion measure was the performance on the second task. The primary purpose of the study was to measure possible effects of transfer of training; the latter would be indicated by the magnitude of the interaction effects.

The analysis of variance in the Gordon study had the following form:

Source	df	MS	F
A (first task)	3	.21	1.41
B (second task)	3	4.70	32.38
AB	9	.14	3.82
Within cell	144	.038	

The analysis given above represents only part of the analysis; separate analyses were made for different stages in the learning of the second task. Since both factors A and B were considered random, the F ratios had the following form:

$$F = \frac{MS_a}{MS_{ab}}, \quad F = \frac{MS_b}{MS_{ab}}, \quad F = \frac{MS_{ab}}{MS_{w, cell}}.$$

These are the F ratios which would be obtained if one were to derive the expected values of the mean squares. The critical value for a .05-level test on the interaction is $F_{.95}(9,144) = 1.94$. (In the opinion of the author the factors in this kind of experiment should be considered fixed rather than random; the study did not indicate the randomization procedure whereby the tasks used in the experiment were obtained. Rather, the indications were that the tasks were deliberately chosen to meet specified objectives.)

A $6 \times 4 \times 3$ factorial having only one observation per cell is reported

by Aborn et al. (1959). These writers were interested in studying the effects of contextual constraints upon recognition of words in sentences. The observation in each of the cells was a mean. A square-root transformation was used before the analysis was made. Entries in three of the cells were missing and were estimated from the data in the remaining cells. All the interactions were pooled into a single error term. The analysis of variance had the following form:

Source	df	MS	F
<i>A</i> (class of word)	5	4.44	26.28
<i>B</i> (position)	3	.47	2.76
<i>C</i> (sentence length)	2	.54	3.22
Pooled interactions	$61 - 3 = 58$.17	

The pooled interaction term would have 61 degrees of freedom if there were no missing data; since 3 of the 72 cell entries were missing, degrees of freedom for the resulting pooled interaction are $61 - 3 = 58$. The separate interaction terms have the following degrees of freedom:

$$AB: 5 \times 3 = 15$$

$$AC: 5 \times 2 = 10$$

$$BC: 3 \times 2 = 6$$

$$ABC: 5 \times 3 \times 2 - 3 = 27$$

Bamford and Ritchie (1958) report an experiment which may be represented as follows:

Subject	a_1				a_2			
	b_1	b_2	b_3	b_4	b_1	b_2	b_3	b_4
1								
2								
.								
.								
9								

The levels of factor *A* represent control and experimental conditions for turn indicators on an airplane instrument panel. The levels of factor *B* represent successive trials under each condition. Each of the nine subjects was observed under each of the levels of factor *A* as well as each of the levels of factor *B*. The order in which trials were made under the level of factor *A* was counterbalanced.

This design, as well as the others which follow in this section, actually falls in the repeated-measures category. The usual tests on main effects and

interactions are valid only under a set of highly restrictive assumptions on covariance matrices. Checks on these assumptions have, for the most part, been ignored by experimenters in the behavioral sciences. The usual tests tend to err on the side of yielding too many significant results (positive bias) when homogeneity assumptions on covariance matrices are not met. A test procedure which avoids this kind of bias is discussed in Sec. 7.2. In experiments involving a learning or practice effect, homogeneity conditions required for the usual F tests to be valid are generally not present.

In form, this plan may be considered as a special case of a $2 \times 4 \times 9$ factorial experiment. The subject factor is considered random; factors A and B are considered fixed. Assuming that all interactions with the subject factor may be pooled, i.e., that all interactions with the subject factor are independent estimates of experimental error, the analysis of variance has the following form:

Source	df	MS	F
Subjects	8	11.83	
A	1	20.77	5.82
B	3	11.42	3.19
AB	3	9.14	2.56
Pooled error	56	3.57	

Gerathewohl et al. (1957) report a study having the form of a $2 \times 9 \times 9$ factorial experiment. The purpose of this experiment was to study the effects of speed and direction of rotation upon the pattern of circular eye

Table 6.13-1 Summary of Analysis of Variance

Source	df	MS	F
C (subjects)	8	30.80	
A (direction)	1	47.22	3.41
AC	8	13.84	
B (speed)	8	85.66	17.8
BC	64	4.80	
AB	8	1.75	.63
ABC	64	2.77	
Total	161		

movements. Factor A represented the direction of rotation; factor B represented the speed of the rotation. Factor C represented a subject factor. The order in which subjects were observed under the levels of factors A and B was the same for all subjects. The subject factor was considered random, but factors A and B were considered fixed. The expected values of the mean squares for this design are analogous to those obtained in Table 6.5-3 when homogeneity assumptions discussed in Sec.

7.2 are met. The analysis of variance reported by these workers is summarized (with slight modification) in Table 6.13-1.

Inspection of the interactions with the subject factor indicates that no pooling is possible. Hence the F ratios have the following form:

$$F = \frac{MS_a}{MS_{ac}}, \quad F = \frac{MS_b}{MS_{bc}}, \quad F = \frac{MS_{ab}}{MS_{abc}}.$$

Jerison (1959) reports a study which is a special case of a $2 \times 3 \times 4 \times 9$ factorial experiment. The purpose of this experiment was to study the effects of noise on human performance. The levels of factor A represented

Table 6.13-2 Summary of Analysis of Variance

Source	df	MS	F
Subjects	8	6544.90	
A (noise conditions)	1	8490.08	2.93
$A \times$ subjects	8	2900.09	
B (periods)	3	479.67	6.32
$B \times$ subjects	24	75.89	
C (clocks)	2	489.31	1.24
$C \times$ subjects	16	396.08	
AB	3	600.47	3.48
$AB \times$ subjects	24	172.60	
AC	2	280.52	1.18
$AC \times$ subjects	16	238.36	
BC	6	138.63	1.32
$BC \times$ subjects	48	105.09	
ABC	6	253.10	2.07
$ABC \times$ subjects	48	122.13	
Total	215		

control and experimental noise conditions. The levels of factor B represented four successive periods of time under each of the conditions of factor A . The levels of factor C represented three clocks monitored by the subjects during the course of the experiment. The last factor represented a subject factor. The criterion was an accuracy score on each of the clocks. All factors except the subject factor were considered to be fixed. The analysis of variance is summarized in Table 6.13-2. In each case the denominator for an F ratio is the corresponding interaction with the subject factor.

6.14 Unequal Cell Frequencies—Least-squares Solution

The data in Table 6.14-1 will be used to illustrate the computational procedures associated with the least-squares estimates of the sums of squares.

The rationale underlying these procedures was discussed in Sec. 5.23. Definitions of the basic symbols are also given in the latter section.

These data have been taken from Anderson and Bancroft (1952, p. 243). The levels of factor A represent sex of animal; levels of factor B represent

Table 6.14-1 Numerical Example with Unequal Cell Frequencies
($p = 2$)

Cell frequencies:					
	b_1	b_2	b_3	b_4	Total
(i) a_1	$n_{11} = 21$	$n_{12} = 15$	$n_{13} = 12$	$n_{14} = 7$	$n_{1.} = 55$
a_2	$n_{21} = 27$	$n_{22} = 25$	$n_{23} = 23$	$n_{24} = 19$	$n_{2.} = 94$
	$n_{.1} = 48$	$n_{.2} = 40$	$n_{.3} = 35$	$n_{.4} = 26$	$n_{..} = 149$
Cell totals:					
	b_1	b_2	b_3	b_4	
(ii) a_1	$AB_{11} = 3716$	$AB_{12} = 2422$	$AB_{13} = 1868$	$AB_{14} = 1197$	$A_1 = 9203$
a_2	$AB_{21} = 2957$	$AB_{22} = 2852$	$AB_{23} = 2496$	$AB_{24} = 2029$	$A_2 = 10334$
	$B_1 = 6673$	$B_2 = 5274$	$B_3 = 4364$	$B_4 = 3226$	$G = 19537$
(iii)					
	$(1) = G^2/n_{..}$				$= 2,561,707$
	$(2) = \sum X^2$				$= 2,738,543$
	$(3) = \sum(A_i^2/n_{i.}) = 1,539,913 + 1,136,080$				$= 2,675,993$
	$(4) = \sum(B_j^2/n_{.j})$				$= 2,567,463$
	$(5) = \sum \sum (AB_{ij}^2/n_{ij})$				$= 2,680,848$

four successive generations of animals. In this type of experiment, the cell frequencies are, in a real sense, an integral part of the design. Hence a least-squares analysis is more appropriate than an unweighted-means

Table 6.14-2 Direct Computation of $SS_{ab(adj)}$

Cell means:					
	b_1	b_2	b_3	b_4	
a_1	176.95	161.47	155.67	171.00	
a_2	109.52	114.08	108.52	106.79	
d_j	67.43	47.39	47.15	64.21	
w_j	11.81	9.38	7.88	5.12	$\sum w_j = 34.19$
$w_j d_j$	796.35	444.52	371.54	328.76	$\sum w_j d_j = 1941.17$

analysis. In part iii, with the exception of symbol (2), data from which the numerical values of the symbols are computed are given in parts i and ii. The raw data for symbol (2) are not given.

When one of the factors has only two levels, the simplest approach (see Rao, 1952, pp. 95-100) is to obtain $SS_{ab(adj)}$ directly from part i of Table

6.14-1 and from Table 6.14-2. In the latter table,

$$d_j = \overline{AB}_{1j} - \overline{AB}_{2j} \quad \text{and} \quad w_j = \frac{n_{1j}n_{2j}}{n_{1j} + n_{2j}}.$$

For example, $w_1 = \frac{(21)(27)}{48} = 11.81.$

The adjusted sum of squares due to interaction is given by

$$SS_{ab(\text{adj})} = \sum w_j d_j^2 - \frac{(\sum w_j d_j)^2}{\sum w_j}.$$

The first term on the right is most readily obtained from the terms $w_j d_j$, that is,

$$\begin{aligned} \sum w_j d_j^2 &= d_1(w_1 d_1) + d_2(w_2 d_2) + \cdots + d_4(w_4 d_4) \\ &= (67.43)(796.35) + (47.39)(444.52) + \cdots + (64.21)(328.76) \\ &= 113,391. \end{aligned}$$

Thus, $SS_{ab(\text{adj})} = 113,391 - \frac{(1941.17)^2}{34.19} = 3180.$

Table 6.14-3 Summary of Analysis of Variance (Least Squares)

Source of variation		SS	df	MS
<i>A</i>	Sex	110,205	1	110,205
<i>B</i>	Generations	1,675	3	555
<i>AB</i>		3,180	3	1,060
	Error	57,695	141	409.2

From part iii of Table 6.14-1, the unadjusted sums of squares are

$$SS_{\text{cells}} = (5) - (1) = 119,141,$$

$$SS_a = (3) - (1) = 114,286,$$

$$SS_b = (4) - (1) = 5756.$$

The adjusted sum of squares due to factor *A* is

$$SS_{a(\text{adj})} = SS_{\text{cells}} - SS_{ab(\text{adj})} - SS_b = 110,205.$$

Similarly, $SS_{b(\text{adj})} = SS_{\text{cells}} - SS_{ab(\text{adj})} - SS_a = 1675.$

From part iii of Table 6.14-1 one obtains

$$SS_{\text{error}} = (2) - (5) = 57,695.$$

A summary of the analysis of variance is given in Table 6.14-3. Tests follow the same pattern as that of the usual factorial experiment in which model I is appropriate.

The direct computation of $SS_{b(\text{adj})}$ will be illustrated by data in Table 6.14-4. This procedure, or a modified version thereof, is the one generally followed in a $p \times q$ factorial experiment when the smaller of p and q is

Table 6.14-4 Computation of $SS_{b(\text{adj})}$

Cell frequencies:

		b_1	b_2	b_3	b_4	Total	
(i)	a_1	21	15	12	7	55	$A_1 = 9203$
	a_2	27	25	23	19	94	$A_2 = 10334$
		48	40	35	26	149	

		b_1	b_2	b_3	b_4	B'
(ii)	b_1	32.2265	-12.9081	-11.1882	-8.1302	190.85
	b_2		29.2602	-9.3897	-6.9623	15.69
	b_3			26.7542	-6.1762	-172.46
	b_4				21.2687	-34.08

(iii)	b'_1	32.2265	-12.9081	-11.1882	-8.1302	190.85
	b''_1	1.0000	-.4005	-.3472	-.2523	5.9219
	b'_2		24.0905	-13.8706	-10.2184	92.1254
	b''_2		1.0000	-.5758	-.4242	3.8241
	b'_3			14.8830	-14.8828	-53.1511
	b''_3			1.0000	-1.0000	-3.5713

$SS_{b(\text{adj})} = (190.85)(5.9219) + (92.1254)(3.8241) + (-53.1511)(-3.5713)$
 $= 1672.34$

greater than 2. The entries in part i are the cell frequencies. The entries in part ii to the left of the B'_j column are the n'_{ij} . For example,

$$n'_{11} = n_{1.} - \frac{n_{11}^2}{n_{1.}} - \frac{n_{12}^2}{n_{2.}} = 48 - \frac{(21)^2}{55} - \frac{(27)^2}{94} = 32.2265,$$

$$n'_{12} = -\left[\frac{n_{11}n_{12}}{n_{1.}} + \frac{n_{21}n_{22}}{n_{2.}}\right] = -\left[\frac{(21)(15)}{55} + \frac{(27)(25)}{94}\right] = -12.9081,$$

$$n'_{13} = -\left[\frac{n_{11}n_{13}}{n_{1.}} + \frac{n_{21}n_{23}}{n_{2.}}\right] = -\left[\frac{(21)(12)}{55} + \frac{(27)(23)}{94}\right] = -11.1882.$$

As a numerical check,

$$\sum_i n'_{ij} = 0.$$

For example, $\sum_i n_{i2} = -12.9081 + 29.2602 - 9.3897 - 6.9623 = 0$.

The entries in column B'_j of part iii are the adjusted totals for the levels of factor B ; for example,

$$B'_1 = B_1 - \left[\frac{n_{11}A_1}{n_{1.}} + \frac{n_{21}A_2}{n_{2.}} \right] = 6673 - \left[\frac{(21)(9203)}{55} + \frac{(27)(10,334)}{94} \right] = 190.85,$$

$$B'_2 = B_2 - \left[\frac{n_{12}A_1}{n_{1.}} + \frac{n_{22}A_2}{n_{2.}} \right] = 5274 - \left[\frac{(15)(9203)}{55} + \frac{(25)(10,334)}{94} \right] = 15.69.$$

As a check on the numerical work $\Sigma B'_j = 0$.

Part iii represents what is known as the abbreviated Doolittle algorithm. Row b'_1 is identical to row b_1 in part ii. Entries in row b'_1 are obtained by dividing each entry in row b_1 by the first entry in the row, which is 32.2265.

Entries in row b'_2 are obtained from the following relation:

$$b'_{2j} = b_{2j} - b''_{12}b'_{1j}, \quad j \geq 2,$$

where b'_{2j} is the entry in row b'_2 , column b_j , and b_{2j} is the entry in row b_2 , column b_j . For example,

$$b'_{22} = b_{22} - b''_{12}b'_{12} = 29.2605 - (-.4005)(-12.9081),$$

$$b'_{23} = b_{23} - b''_{12}b'_{13} = -9.3897 - (-.4005)(-11.1882).$$

Entries in row b'_2 are obtained by dividing each of the entries in row b'_2 by the entry b'_{22} , which in this case is 24.0905.

Entries in row b'_3 are obtained from the following relation:

$$b'_{3j} = b_{3j} - b''_{13}b'_{1j} - b''_{23}b'_{2j}, \quad j \geq 3.$$

For example, $b'_{34} = b_{34} - b''_{13}b'_{14} - b''_{23}b'_{24}$

$$= -6.9623 - (-.3472)(-8.1302) - (-.5758)(-10.2184)$$

$$= -14.8828.$$

Entries in row b'_3 are obtained by dividing each entry in row b'_3 by b'_{33} , which is 14.8830.

In general, an entry in row b'_i is given by the relation

$$b'_{ij} = b_{ij} - b''_{1i}b'_{1j} - b''_{2i}b'_{2j} - \cdots - b''_{(i-1)i}b'_{(i-1)j},$$

and an entry in row b''_i is the corresponding entry in row b'_i divided by b'_{ii} .

When the abbreviated Doolittle algorithm is used in this manner, there is a simple check on the work: the sum of the entries in row b''_i , excluding the entry in column B'_j , will be zero. If there are q levels of factor B , the last row in the algorithm will be b''_{q-1} .

The numerical value of $SS_{b(\text{adj})}$ is obtained from the B'_j column of the abbreviated Doolittle algorithm.

$$SS_{b(\text{adj})} = \sum_i b''_{iB}b'_{iB}.$$

Having $SS_{b(adj)}$, the adjusted interaction sum of squares is given by

$$SS_{ab(adj)} = SS_{cells} - SS_a - SS_{b(adj)},$$

and the adjusted sum of squares due to the main effect of factor A is

$$\begin{aligned} SS_{a(adj)} &= SS_a + SS_{b(adj)} - SS_b \\ &= SS_{cells} - SS_{ab(adj)} - SS_b. \end{aligned}$$

By way of contrast the unweighted-means analysis is given in Table 6.14-5. In this case the two approaches lead to comparable end products.

Table 6.14-5 Unweighted-means Analysis

Source of variation		SS	df	MS
<i>A</i>	Sex	155,760	1	155,760
<i>B</i>	Generations	1,965	3	655
<i>AB</i>		2,721	3	907
	Error	57,695	141	409.2

Although the abbreviated Doolittle algorithm is rather widely used, a somewhat simpler computational scheme is that known as the Dwyer square-root algorithm. The latter may also be used to solve the normal equations in the general linear-regression problem. The data in Table 6.14-6 will be used to illustrate the Dwyer algorithm in obtaining $SS_{b(adj)}$. The starting point is the data given in part ii of Table 6.14-4.

Table 6.14-6 Computation of $SS_{b(adj)}$
(Dwyer square-root algorithm)

	b_1	b_2	b_3	b_4	B'
b_1	32.2265	-12.9081	-11.1882	-8.1302	190.85
b_2		29.2602	-9.3897	-6.9623	15.69
b_3			26.7542	-6.1762	-172.46
b_4				21.2687	-34.08
b'_1	5.6769	-2.2739	-1.9708	-1.4322	33.6187
b'_2		4.9081	-2.8262	-2.0821	18.7721
b_3			3.8578	-3.8579	-13.7774
$SS_{b(adj)} = (33.6187)^2 + (18.7721)^2 + (-13.7774)^2 = 1672.43$					

The entries in row b'_1 are

$$b'_{11} = \sqrt{b_{11}}, \quad b'_{1j} = \frac{b_{1j}}{b'_{11}}, \quad j > 1.$$

For example,

$$b'_{11} = \sqrt{32.2265} = 5.6769; \quad b'_{12} = \frac{-12.9081}{5.6769}.$$

The entries in the row b'_2 are defined as follows:

$$b'_{22} = \sqrt{b_{22} - b_{12}^2}, \quad b'_{2j} = \frac{b_{2j} - b'_{12}b'_{1j}}{b'_{22}}, \quad j > 2.$$

For example, $b'_{22} = \sqrt{29.2602 - (-2.2739)^2} = 4.9081$

$$b'_{23} = \frac{-9.3897 - (-2.2739)(-1.9708)}{4.9081} = -2.8262.$$

The entries in row b'_3 are defined as follows:

$$b'_{33} = \sqrt{b_{33} - b_{13}^2 - b_{23}^2};$$

$$b'_{3j} = \frac{b_{3j} - b'_{13}b'_{1j} - b'_{23}b'_{2j}}{b'_{33}}, \quad j > 3.$$

In general, the entries in row b'_i are defined as follows:

$$b'_{ii} = \sqrt{b_{ii} - \sum_k b_{ki}^2}, \quad \text{where } k = 1, 2, \dots, i-1;$$

$$b'_{ij} = \frac{b_{ij} - \sum_k b_{ki}b_{kj}}{b'_{ii}}, \quad \text{where } j > i.$$

Entries under the B'_j column are considered in the same category as any other column.

To obtain $SS_{b(\text{adj})}$, the algorithm terminates after $q-1$ rows are complete. The adjusted sum of squares for factor B is obtained from the B'_j column of the algorithm from the relationship shown in the last line of Table 6.14-6. As a check on the computational work, the sum of the entries in each b' row (excluding the entry under column B'_j) is zero (within rounding error).

CHAPTER 7

Multifactor Experiments Having Repeated Measures on the Same Elements

7.1 General Purpose

Factorial experiments in which the same experimental unit (generally a subject) is observed under more than one treatment condition require special attention. Experiments of this kind will be referred to as those in which there are repeated measures. Single-factor experiments of this kind were discussed in considerable detail in Chap. 4.

Different approaches have been used in handling the sampling distributions of mean squares which arise in the analysis of experiments. These approaches differ to some extent in the definition of the experimental unit. Under one approach, which assumes a multivariate normal underlying distribution in the population, the experimental errors are considered to be correlated. Under an alternative approach, a nested random factor (a kind of dummy variable) is included in the model to absorb the correlation between the experimental errors. This latter approach has the advantage of permitting a relatively simple derivation of the expected values of the mean squares. The disadvantage of the latter approach is that it tends to mask homogeneity assumptions (specifically, constant correlation between pairs of observations) required to justify the sampling distribution of the final F ratio used in making tests. Both approaches lead to identical F ratios when the homogeneity assumptions are met. A comprehensive discussion of the homogeneity assumptions underlying repeated measure designs will be found in the work of Bargmann (1957).

A two-factor experiment in which there are repeated measures on factor B (that is, each experimental unit is observed under all levels of factor B) may be represented schematically as follows:

	b_1	b_2	b_3
a_1	G_1	G_1	G_1
a_2	G_2	G_2	G_2

The symbol G_1 represents a group of n subjects. The symbol G_2 represents a second group of n subjects. The subjects in G_1 are observed under treatment combinations ab_{11} , ab_{12} , and ab_{13} . Thus the subjects in G_1 are observed under all levels of factor B in the experiment, but only under one level of factor A . The subjects in G_2 are observed under treatment combinations ab_{21} , ab_{22} , and ab_{23} . Thus each subject in G_2 is observed under all levels of factor B in the experiment, but only under one level of factor A , namely, a_2 .

In this kind of experiment, comparisons between treatment combinations at different levels of factor A involve differences between groups as well as differences associated with factor A . On the other hand, comparisons between different levels of factor B at the same level of A do not involve differences between groups. Since measurements included in the latter comparisons are based upon the same elements, main effects associated with such elements tend to cancel. For the latter comparisons, each element serves as its own control with respect to such main effects.

In an experiment of the type represented above, the main effects of factor A are said to be completely confounded with differences between groups. On the other hand, the main effects of factor B as well as the AB interaction will be shown to be free of such confounding. Tests on B and AB will generally be considerably more sensitive than tests on the main effects of factor A . Where no confounding with the group factor is present, there are fewer uncontrolled sources of error variance. The smaller the error variance, the more sensitive (powerful) the test.

By using the approach which assumes that the errors are correlated, the expected value of the variance due to the main effects of factor A has the general form

$$E(MS_a) = \sigma_e^2[1 + (q - 1)\rho] + nq\sigma_a^2$$

where ρ is the (constant) correlation between pairs of observations on the same element and q is the number of levels of factor B . The expected value of the mean square appropriate for the denominator of an F ratio used in testing the variance due to the main effect of factor A has the form

$$\sigma_e^2[1 + (q - 1)\rho].$$

For sources of variation which are not confounded with the main effects due to differences between groups, the denominator of an F ratio has the expected value

$$\sigma_e^2(1 - \rho).$$

Thus, if correlation between pairs of measurements is positive and constant, the latter experimental error will be smaller than the former. Hence

the greater sensitivity (i.e., the greater power) of tests using the latter error term.

In terms of the alternative approach, which postulates the existence of a nested random factor, the expected value for the denominator of an F ratio for a test on the main effects of factor A has the form

$$\sigma_e^2 + q\sigma_{\pi}^2,$$

where σ_{π}^2 is the variation due to the main effects of subjects within the groups. The expected value for the denominator for tests on effects which are not confounded with main effects due to subjects within the groups has the form

$$\sigma_e^2 + \sigma_{\beta\pi}^2,$$

where $\sigma_{\beta\pi}^2$ is the interaction between the subject and treatment factors. The magnitude of $\sigma_{\beta\pi}^2$ is generally considerably smaller than σ_{π}^2 . (It should be noted that σ_e^2 includes more sources of variance than does σ_{π}^2 .)

Repeated measures on the same elements may arise in different ways. In experiments designed to study rates of learning as a function of treatment effects, repeated measures on the same subject are a necessary part of the design. Further, the order in which the observations are made is dictated by the experimental variables. On the other hand, in experiments designed to evaluate the joint effect of two or more treatments the experimenter may have his option as to whether or not the same elements are observed under more than one treatment combination. Further, the order in which the elements appear under the treatment combinations may also be under the control of the experimenter. The utility of designs calling for repeated measures is limited where carry-over effects are likely to confound results. In some cases such effects may be controlled by counterbalancing the order in which treatment combinations are given to the elements. (Designs of this kind are discussed in Sec. 10.7.)

Aside from designs having the form of learning experiments, the primary purpose of repeated measures on the same elements is the control that this kind of design provides over individual differences between experimental units. In the area of the behavioral sciences, differences between such units often are quite large relative to differences in treatment effects which the experimenter is trying to evaluate. If there are no carry-over effects, a repeated-measures design in the area of the behavioral sciences is to some degree analogous to a split-plot design in the area of agricultural experimentation. Where additive carry-over effects are present, repeated-measures designs are analogous to crossover designs.

Another (somewhat doubtful) advantage of a repeated-measures design is

in terms of economy of subjects. Using different subjects under each of the treatment combinations in a factorial experiment has the *marked advantage* of providing statistically independent estimates of treatment effects from all cells in the experiment. Increasing the number of statistically independent observations is very likely to be the best way of increasing the precision of estimators. By having each subject serve as his own control, the experimenter attempts to work with a smaller sample size. However, the simple additive model underlying the usual analysis for the case of repeated measures may not be an adequate representation of the experimental phenomena. A more inclusive multivariate regression model is often required to represent fully the underlying experimental variables.

In sections that follow, special cases of factorial designs having repeated measures will be illustrated. The expected values of the mean square for these special cases will be given. The latter are obtained by means of the methods given by Cornfield and Tukey (1956). The validity of the tests set up on the basis of such expected values rests upon assumptions about the form of the variance-covariance matrix associated with the joint multivariate normal distribution of the variables in the model. Unless the nature of the experimental variables dictates the order in which treatments are administered to subjects, it will be assumed that the order of administration is randomized independently for each of the subjects. Further, it will be assumed that the n elements in a group are a random sample from a specified population of elements.

A strong word of warning is required in connection with order (or sequence) effects. Practice, fatigue, transfer of training, the effects of an immediately preceding success or failure illustrate what fall in the latter category. If such effects exist, randomizing or counterbalancing does not remove them; rather, such procedures completely entangle the latter with treatment effects. There is some chance that sequence effects will balance out—they generally will if a simple additive model is realistic. However, in experiments (other than those which primarily are concerned with learning or carry-over effects) where the sequence effects are likely to be marked and where primary interest lies in evaluating the effect of individual treatments in the absence of possible sequence effects, a repeated-measures design is to be avoided.

In cases where sequence effects are likely to be small relative to the treatment effects, repeated-measures designs can be used. Counterbalancing or randomizing order of administration in this case tends to prevent sequence effects from being completely confounded with one or just a selected few of the treatments. Instead such sequence effects are spread over all the treatment effects. Admittedly such sequence effects may serve to mask treatment effects; however, the potential advantages can outweigh the potential disadvantages.

7.2 Two-factor Experiment with Repeated Measures on One Factor

This kind of experiment was illustrated in the last section. The general case may be represented as follows:

	b_1	\cdots	b_j	\cdots	b_q
a_1	G_1	\cdots	G_1	\cdots	G_1
\vdots					
a_i	G_i	\cdots	G_i	\cdots	G_i
\vdots					
a_p	G_p	\cdots	G_p	\cdots	G_p

Each G represents a random sample of size n from a common population of subjects. Each of the subjects in G_i is observed under q different treatment combinations, all of these treatment combinations involving factor A at level a_i . The actual observations on the subjects within group i may be represented as follows:

	Subject	b_1	\cdots	b_j	\cdots	b_q
	1	X_{i11}	\cdots	X_{ij1}	\cdots	X_{iq1}
	\vdots					
a_i	k	X_{i1k}	\cdots	X_{ijk}	\cdots	X_{iqk}
	\vdots					
	n	X_{i1n}	\cdots	X_{ijn}	\cdots	X_{iqn}

The symbol X_{ijk} denotes a measurement on subject k in G_i under treatment combination ab_{ij} . A more complete notation for subject k would be $k(i)$; the latter notation distinguishes this subject from subject k in some other group. Similarly a more complete notation for X_{ijk} would be $X_{ijk(i)}$. The latter notation is rather cumbersome; in cases in which there is no ambiguity, the symbol X_{ijk} will be used to indicate an observation on subject k in G_i made under treatment combination ab_{ij} .

For special cases the notation for a subject may be made more specific. For example, consider the case in which $p = 2$, $q = 3$, and $n = 2$. The

experimental data may be represented as follows:

	Subject	b_1	b_2	b_3
a_1	1	X_{111}	X_{121}	X_{131}
	2	X_{112}	X_{122}	X_{132}
a_2	3	X_{213}	X_{223}	X_{233}
	4	X_{214}	X_{224}	X_{234}

In this notation scheme subject 3 is the first subject in G_2 , and subject 4 is the second subject in G_2 .

The linear model upon which the analysis will be based has the following form:

$$X_{ijk} = \mu + \alpha_i + \pi_{k(i)} + \beta_j + \alpha\beta_{ij} + \beta\pi_{jk(i)} + \varepsilon_{k(ij)}.$$

The notation $\pi_{k(i)}$ indicates that the effect of subject k is nested under level a_i . Note that the linear model does not include any carry-over effects. The analysis of variance for this kind of experiment takes the form given in Table 7.2-1. The expected values shown in this table assume that A and B are fixed factors.

Table 7.2-1 Summary of Analysis of Variance

Source of variation	df	E(MS)
Between subjects	$np - 1$	
A	$p - 1$	$\sigma_\varepsilon^2 + q\sigma_\pi^2 + nq\sigma_\alpha^2$
Subjects within groups	$p(n - 1)$	$\sigma_\varepsilon^2 + q\sigma_\pi^2$
Within subjects	$np(q - 1)$	
B	$q - 1$	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2 + np\sigma_\beta^2$
AB	$(p - 1)(q - 1)$	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2 + n\sigma_{\alpha\beta}^2$
$B \times$ subjects within groups	$p(n - 1)(q - 1)$	$\sigma_\varepsilon^2 + \sigma_{\beta\pi}^2$

The manner in which the total variation is partitioned in this table is quite similar to that used in a $p \times q$ factorial experiment in which there are no repeated measures. A comparison of the two partitions is shown in Table 7.2-2. It will be noted that partition of the between-cell variation is identical. However, in an experiment having repeated measures, the within-cell variation is divided into two orthogonal (nonoverlapping) parts. One part is a function of experimental error plus the main effects of subjects within groups, i.e., individual differences. The other part is a function of experimental error and $B \times$ subject-within-group interaction. If the latter interaction is negligible, then the second part of the within-cell variation is a function solely of experimental error.

Table 7.2-2 Comparison of Partitions

$p \times q$ factorial (no repeated measures)		$p \times q$ factorial (repeated measures on factor B)	
Total	$npq - 1$	Total	$npq - 1$
Between cells	$pq - 1$	Between cells	$pq - 1$
A $p - 1$		A $p - 1$	
B $q - 1$		B $q - 1$	
AB $(p - 1)(q - 1)$		AB $(p - 1)(q - 1)$	
Within cells	$pq(n - 1)$	Within cells	$pq(n - 1)$
		Subjects within groups	$p(n - 1)$
		$B \times$ subjects within groups	$p(n - 1)(q - 1)$

Appropriate denominators for F ratios to be used in making statistical tests are indicated by the expected values of the mean squares. Thus, to test the hypothesis that $\sigma_{\alpha}^2 = 0$,

$$F = \frac{MS_a}{MS_{\text{subj w. groups}}}$$

The mean square in the denominator of the above F ratio is sometimes designated $MS_{\text{error (between)}}$. To test the hypothesis that $\sigma_{\beta}^2 = 0$,

$$F = \frac{MS_b}{MS_{B \times \text{subj w. groups}}}$$

To test the hypothesis that $\sigma_{\alpha\beta}^2 = 0$, the appropriate F ratio is

$$F = \frac{MS_{ab}}{MS_{B \times \text{subj w. groups}}}$$

The mean square in the denominator of the last two F ratios is sometimes called $MS_{\text{error (within)}}$ since it forms the denominator of F ratios used in testing effects which can be classified as part of the within-subject variation.

The mean squares used in the denominators of the above F ratios represent a pooling of different sources of variation. The variation due to subjects within groups is the sum of the following sources of variation:

Source	df
Subjects within groups	$p(n - 1)$
Subjects within G_1	$n - 1$
Subjects within G_2	$n - 1$
.....
Subjects within G_p	$n - 1$

One of the assumptions required in order that the F ratio actually follow an F distribution is that these sources of variation be homogeneous. A partial check on this assumption may be made through use of the statistic

$$F_{\max} = \frac{\text{maximum } (SS_{\text{subj w. } G_i})}{\text{minimum } (SS_{\text{subj w. } G_i})},$$

i.e., the ratio of the largest of these sources of variation to the smallest. The critical value for this statistic in a test having level of significance equal to α is

$$F_{\max(1-\alpha)}(p, n-1).$$

These critical values are given in Table B.7.

The variation due to $B \times$ subjects within groups represents a pooling of the following sources:

Source	df
$B \times$ subjects within groups	$p(n-1)(q-1)$
$B \times$ subjects within G_1	$(n-1)(q-1)$
$B \times$ subjects within G_2	$(n-1)(q-1)$
.....
$B \times$ subjects within G_p	$(n-1)(q-1)$

A test on the homogeneity of these sources is given by

$$F_{\max} = \frac{\text{maximum } (SS_{B \times \text{subj w. } G_i})}{\text{minimum } (SS_{B \times \text{subj w. } G_i})}.$$

The critical value for this test is

$$F_{\max(1-\alpha)}[p, (n-1)(q-1)].$$

If the scale of measurement for the original criterion data does not satisfy these homogeneity assumptions, a transformation may often be found which will satisfy these assumptions. Indications are, however, that the F tests given above are robust with respect to minor violations of these assumptions [see Box (1954)].

For the sampling distribution of the F ratio for within-subject effects to be the F distribution with the usual degrees of freedom requires additional assumptions about the pattern of elements in $q \times q$ covariance matrices. (The latter covariance matrices appear in a later section of this chapter.) A method for making this check is illustrated in the work of Greenhouse and Geisser (1959); a numerical example is given in Sec. 7.7. These workers also provide a simple approximation procedure which avoids assumptions about equal covariances in the pooled variance-covariance matrix. The approximation procedure given by these writers errs on the side of making the critical value somewhat larger than should be the case. (Hence the test procedure is negatively biased, since it will yield too few "significant" decisions.)

In the Greenhouse and Geisser procedure, the F ratios given above are used, but the degrees of freedom used in finding the critical values are adjusted. To test the hypothesis that $\sigma_{\beta}^2 = 0$, the critical value for the F ratio is

$$F_{1-\alpha}[1, p(n-1)] \quad \text{instead of} \quad F_{1-\alpha}[(q-1), p(n-1)(q-1)].$$

To test the hypothesis that $\sigma_{\alpha\beta}^2 = 0$, the critical value is

$$F_{1-\alpha}[(p-1), p(n-1)] \quad \text{instead of} \quad F_{1-\alpha}[(p-1)(q-1), p(n-1)(q-1)].$$

Use of these modified critical values is recommended if there is reason to doubt that the covariances between pairs of observations are constant for all levels of factor B . Transformations on the scale of measurement which yield homogeneity of the error terms which are pooled do not necessarily yield the required homogeneity conditions on the covariances.

Computational Procedures. To illustrate the computational procedures for this plan, consider a factorial experiment in which the levels of factor A

Table 7.2-3 Numerical Example

Observed data:

(i)

	Subject	b_1	b_2	b_3	b_4	Total
a_1	1	0	0	5	3	8 = P_1
	2	3	1	5	4	13 = P_2
	3	4	3	6	2	15 = P_3
a_2	4	4	2	7	8	21 = P_4
	5	5	4	6	6	21 = P_5
	6	7	5	8	9	29 = P_6
						107 = G

AB summary table:

(ii)

	b_1	b_2	b_3	b_4	Total
a_1	7	4	16	9	36
a_2	16	11	21	23	71
	23	15	37	32	107

Computational symbols:

(iii)

$$\begin{aligned}
 (1) &= G^2/npq = (107)^2/3(2)(4) = 477.04 \\
 (2) &= \Sigma X^2 = 0^2 + 0^2 + 5^2 + \cdots + 8^2 + 9^2 = 615 \\
 (3) &= (\Sigma A_i^2)/nq = (36^2 + 71^2)/3(4) = 528.08 \\
 (4) &= (\Sigma B_j^2)/np = (23^2 + 15^2 + 37^2 + 32^2)/3(2) = 524.50 \\
 (5) &= [\Sigma (AB_{ij})^2]/n = (7^2 + 4^2 + \cdots + 21^2 + 23^2)/3 = 583.00 \\
 (6) &= (\Sigma P_k^2)/q = (8^2 + 13^2 + \cdots + 21^2 + 29^2)/4 = 545.25
 \end{aligned}$$

are two methods for calibrating dials and the levels of B are four shapes for the dials. Suppose that the data obtained are those given in part i of Table 7.2-3. Entries are accuracy scores on a series of trials on each of the dials. Thus, for this experiment $p = 2$, $q = 4$, and $n = 3$. The order in which subjects are observed under the dials is randomized independently.

From the data in part i, the AB summary table given in part ii is readily obtained. Computational symbols are defined and computed in part iii. The only symbol that does not occur in a $p \times q$ factorial experiment without repeated measures is (6). This latter symbol involves P_k , which is the sum of the q observations made on subject k . These sums are given at the right of part i. In each case the divisor in a computational symbol is the number of observations summed to obtain an entry which is squared in the numerator.

Table 7.2-4 Analysis of Variance for Numerical Example

Source of variation	Computational formula	SS	df	MS	F
Between subjects	(6) - (1) = 68.21	5			
A (calibration)	(3) - (1) = 51.04	1		51.04	11.90
Subjects within groups	(6) - (3) = 17.17	4		4.29	
Within subjects	(2) - (6) = 69.75	18			
B (shape)	(4) - (1) = 47.46	3		15.82	12.76
AB	(5) - (3) - (4) + (1) = 7.46	3		2.49	2.01
$B \times$ subjects within groups	(2) - (5) - (6) + (3) = 14.83	12		1.24	

The analysis of variance is summarized in Table 7.2-4. In terms of means,

$$SS_{\text{subj w. groups}} = q \sum_k (\bar{P}_{k(i)} - \bar{A}_i)^2,$$

$$SS_{B \times \text{subj w. groups}} = \sum_k \sum_j \sum_i (X_{ijk} - \bar{P}_{k(i)} - \bar{B}_j + \bar{A}_i)^2.$$

The computational formulas given in Table 7.2-4 are equivalent to these operations on the means.

The differential sensitivity for tests on between- and within-subject effects should be noted. The denominator of the F ratio for the between-subject effects is 4.29, whereas the denominator of the F ratios for the within-subject effects is 1.24. It is not unusual in this kind of plan to find the ratio of these two denominators to be as large as 10:1. If the level of significance for tests is set at .05, one rejects the hypothesis that $\sigma_\alpha^2 = 0$, since $F_{.95}(1,4) = 7.71$; the hypothesis that $\sigma_\beta^2 = 0$ is also rejected, since $F_{.95}(3,12) = 3.49$. However, the experimental data do not contradict the hypothesis that $\sigma_{\alpha\beta}^2 = 0$.

If the experimenter has reason (often there is reason) to question the homogeneity of the covariances in the underlying population, the critical

values for the within-subject tests are as follows:

Hypothesis	Conservative test	Ordinary test
$\sigma_{\beta}^2 = 0$	$F_{.95}[1,4] = 7.71$	$F_{.95}[3,12] = 3.49$
$\sigma_{\alpha\beta}^2 = 0$	$F_{.95}[1,4] = 7.71$	$F_{.95}[3,12] = 3.49$

Even under the more conservative test (i.e., negatively biased test) the main effects due to the shapes of the dials remain statistically significant.

If the experimenter has reason to question the homogeneity of the parts that are pooled to form the denominators of the F ratios, a check on homogeneity would logically precede the tests. Computational procedures for partitioning the relevant sums of squares are given in Table 7.2-5. A symbol of form $(6a_1)$ has the same general definition as (6), but summations

Table 7.2-5 Partition of Error Terms

(i)	$(6a_1) = (\sum_{a_1} P_k^2)/q = (8^2 + 13^2 + 15^2)/4$	= 114.50
	$(6a_2) = (\sum_{a_2} P_k^2)/q = (21^2 + 21^2 + 29^2)/4$	= 430.75
		(6) = 545.25
	$(3a_1) = (A_1^2)/nq = 36^2/3(4)$	= 108.00
	$(3a_2) = (A_2^2)/nq = 71^2/3(4)$	= 420.08
		(3) = 528.08
	$(5a_1) = [\sum_{a_1} (AB_{ij})^2]/n = [7^2 + 4^2 + 16^2 + 9^2]/3$	= 134.00
	$(5a_2) = [\sum_{a_2} (AB_{ij})^2]/n = [16^2 + 11^2 + 21^2 + 23^2]/3$	= 449.00
		(5) = 583.00
	$(2a_1) = \sum_{a_1} X^2 = 0^2 + 0^2 + \dots + 6^2 + 2^2$	= 150
(ii)	$(2a_2) = \sum_{a_2} X^2 = 4^2 + 2^2 + \dots + 8^2 + 9^2$	= 465
		(2) = 615
	$SS_{\text{subj w. } G_1} = (6a_1) - (3a_1)$	= 6.50
	$SS_{\text{subj w. } G_2} = (6a_2) - (3a_2)$	= 10.67
		17.17
	$SS_{B \times \text{subj w. } G_1} = (2a_1) - (5a_1) - (6a_1) + (3a_1) =$	9.50
	$SS_{B \times \text{subj w. } G_2} = (2a_2) - (5a_2) - (6a_2) + (3a_2) =$	5.33
		14.83

are restricted to level a_1 . The computational procedures for parts which are pooled are given in part ii. As a check on the homogeneity of $SS_{\text{subj w. groups}}$,

$$F_{\max} = \frac{10.67}{6.50} = 1.64.$$

The critical value for a .05-level test here is

$$F_{\max, .95}(2,2) = 39.00.$$

Since the computed F_{\max} statistic does not exceed the critical value, the test does not contradict the hypothesis that the parts are homogeneous. Since each of the parts in this case has only two degrees of freedom, the power of a test of this kind is extremely low.

As a check on the homogeneity of the parts of $SS_{B \times \text{subj w. groups}}$,

$$F_{\max} = \frac{9.80}{5.33} = 1.84.$$

The critical value here is

$$F_{\max, .95}(2, 6) = 5.82.$$

Again the computed value of the statistic does not exceed the critical value. Hence the hypothesis of homogeneity is not contradicted by the experimental data.

Tests on the difference between all possible pairs of means can be made in a manner similar to that given in Sec. 3.8. The procedure is illustrated in Table 7.2-6. The Newman-Keuls method is chosen for this purpose.

Table 7.2-6 Tests on Means Using Newman-Keuls Procedure

	Shapes	b_2	b_1	b_4	b_3
(i)	Ordered means	2.50	3.83	5.33	6.17
		b_2	b_1	b_4	b_3
	Differences	b_2	1.33	2.83	3.67
	between pairs	b_1		1.50	2.34
		b_4			.84
(ii)	$s_B = .46$		$r = 2$	3	4
	$q_{.95}(r, 12):$		3.08	3.77	4.20
	$s_B q_{.95}(r, 12):$		1.42	1.73	1.93
(iii)		b_2	b_1	b_4	b_3
		b_2	—	*	*
		b_1		*	*
		b_4			—

In part i the \bar{B}_j 's are arranged in rank order from low to high. Differences between all possible pairs of ordered means are computed. For example,

$$6.17 - 2.50 = 3.67, \quad 5.33 - 2.50 = 2.83, \quad \text{etc.}$$

In part ii critical values for the ordered differences between pairs are computed. Since the main effect of factor B is a within-subject effect, the standard error of the mean for all observations at a given level of factor B is

$$s_B = \sqrt{\frac{MS_{B \times \text{subj w. groups}}}{np}} = \sqrt{\frac{1.26}{6}} = \sqrt{.207} = .46.$$

The degrees of freedom associated with this standard error are those of $MS_{B \times \text{subj w. groups}}$, which in this case are 12. To obtain the critical value for the difference between two ordered means which are r steps apart in an ordered sequence, one first finds the tabled values for

$$q_{1-\alpha}(r, df_{\text{error}}),$$

where df_{error} represents the degrees of freedom associated with s_B . These values are obtained from Table B.4. For level of significance .05, the relevant values of q are given in part ii. The critical value for an ordered difference between two means r steps apart is

$$s_B q_{1-\alpha}(r, df_{\text{error}}).$$

These critical values also appear in part ii. (For the sequence in which tests on ordered pairs must be made, see Sec. 3.8.)

The pairs of means which can be considered different are indicated in part iii. The mean performance on shape b_3 is statistically different from the mean performance on shapes b_2 and b_1 . The mean performance on shape b_4 is also statistically different from the mean performance on shapes b_2 and b_1 . No other differences are statistically significant at the .05 level for the Newman-Kuels tests.

Tests on all possible ordered differences of the form $\bar{A}_i - \bar{A}_j$ follow the same general pattern. For such tests,

$$s_{\bar{A}} = \sqrt{\frac{MS_{\text{subj w. groups}}}{nq}}.$$

The degrees of freedom associated with $s_{\bar{A}}$ are $p(n - 1)$.

If the AB interaction were significant, tests on simple main effects would be called for, rather than direct tests on main effects. The computation of the variation due to the simple main effect of factors A and B is identical to that of a two-factor factorial experiment which does not have repeated measures. To test the simple main effect of factor B , the F ratio has the form

$$F = \frac{MS_{b \text{ at } a_i}}{MS_{B \times \text{subj w. groups}}}.$$

The denominator of this F ratio is the same as that used in testing the main effects of factor B . The F ratio for the test on the simple main effects of factor A has the form

$$F = \frac{MS_{a \text{ at } b_j}}{MS_{\text{w. cell}}}.$$

The denominator of this last F ratio requires special note—it is not the denominator used in testing the main effects of factor A . For each level of

factor B considered individually, this plan reduces to a single-factor experiment in which there are no repeated measures. In this latter type of experiment $MS_{w, cell}$ is the appropriate denominator for the variation due to the treatment effects.

The within-cell variation is given by

$$SS_{w, cell} = SS_{\text{subj w. groups}} + SS_{B \times \text{subj w. groups}}.$$

Within the context of a repeated-measures design, $SS_{w, cell}$ represents a pooling of what will often be heterogeneous sources of variance. Hence the F test on the simple main effects for factor A , which uses $MS_{w, cell}$ as a denominator, will tend to be biased. However, when the degrees of freedom for the within-cell variation are large (say, greater than 30), the bias will be quite small. The magnitude of the bias depends in part upon the ratio of $MS_{\text{subj w. groups}}$ to $MS_{B \times \text{subj w. groups}}$.

Variation due to the simple main effects is most readily computed from the AB summary table given in part ii of Table 7.2-3. For example,

$$SS_{a \text{ at } b_1} = \frac{7^2 + 16^2}{3} - \frac{23^2}{6} = 13.50;$$

$$MS_{a \text{ at } b_1} = \frac{SS_{a \text{ at } b_1}}{p - 1} = 13.50.$$

The denominator for the appropriate F ratio is

$$MS_{w, cell} = \frac{SS_{w, cell}}{pq(n-1)} = \frac{17.17 + 14.83}{16} = 2.00.$$

$$F = \frac{MS_{a \text{ at } b_1}}{MS_{w, cell}} = 6.75.$$

The degrees of freedom for this F statistic are $[(p - 1), pq(n - 1)]$. The variation due to the simple main effects for factor B at level a_1 is

$$SS_{b \text{ at } a_1} = \frac{7^2 + 4^2 + 16^2 + 9^2}{3} - \frac{36^2}{12} = 26.00;$$

$$MS_{b \text{ at } a_1} = \frac{SS_{b \text{ at } a_1}}{q - 1} = \frac{26.00}{3} = 8.67.$$

A test on the simple main effects of factor B at level a_1 uses the statistic

$$F = \frac{MS_{b \text{ at } a_1}}{MS_{B \times \text{subj w. groups}}} = \frac{8.67}{1.24} = 6.99.$$

The critical value for a test having level of significance .05 is

$$F_{.95}[(q - 1), p(n - 1)(q - 1)] = F_{.95}(3, 12) = 3.49.$$

Hence the experimental data tend to reject the hypothesis that there are no differences in the effects of factor B when all observations are made at level a_1 .

Covariance Matrices Associated with this Design. Consider the following data from part i of Table 7.2-3:

	Subject	b_1	b_2	b_3	b_4
a_1	1	0	0	5	3
	2	3	1	5	4
	3	4	3	6	2
		7	4	16	9

The variance of the observations made under b_1 is

$$\text{var}_{b_1} = \frac{(0^2 + 3^2 + 4^2) - (7^2/3)}{2} = 4.33.$$

Similarly the variance of the observations made under b_2 is

$$\text{var}_{b_2} = \frac{(0^2 + 1^2 + 3^2) - (4^2/3)}{2} = 2.33.$$

The covariance of the observations made under b_1 and b_2 is

$$\text{cov}_{b_1 b_2} = \frac{(0)(0) + (3)(1) + (4)(3) - (7)(4)/3}{2} = 2.83.$$

Similarly, the covariance of the observations under b_1 and b_3 is

$$\text{cov}_{b_1 b_3} = \frac{(0)(5) + (3)(5) + (4)(6) - (7)(16)/3}{2} = .83.$$

The variance-covariance matrix for level a_1 of factor A is given below (only the top half is completed since the bottom half is a duplicate of the top half):

		b_1	b_2	b_3	b_4
B_{a_1}	b_1	4.33			
	b_2	2.83	2.33		
	b_3			.83	-.50
	b_4			.33	1.00

A similar variance-covariance matrix for the data at level a_2 is given below:

		b_1	b_2	b_3	b_4
B_{a_2}	b_1	2.33			
	b_2	2.17	1.00		
	b_3			.50	.33
	b_4			1.00	1.50

In order that the F ratio on within-subject effects follow an F distribution, the population variance-covariance matrices must have the following form:

	b_1	b_2	b_3	b_4
b_1	σ^2	$\rho\sigma^2$	$\rho\sigma^2$	$\rho\sigma^2$
b_2		σ^2	$\rho\sigma^2$	$\rho\sigma^2$
b_3			σ^2	$\rho\sigma^2$
b_4				σ^2

In words, each of the elements along the diagonal is equal to σ^2 , and each of the covariances is equal to $\rho\sigma^2$. It is further assumed that σ^2 and ρ are the same for all levels of factor A . In this context ρ is the product-moment correlation between pairs of measurements on the same subject.

If the underlying population variance-covariance matrices for B_{a_1} and B_{a_2} are equal but not necessarily of the form indicated above, the best estimate of the common underlying population variance-covariance matrix is the pooled sample matrices. The pooled matrix is obtained by averaging corresponding entries in the individual matrices. For the data under consideration, the pooled variance-covariance matrix is given below:

		b_1	b_2	b_3	b_4
B_{pooled}	b_1	3.33	2.50	.92	.33
	b_2		2.33	.67	— .33
	b_3			.67	.50
	b_4				1.67

There are two stages in testing for homogeneity with respect to the variance-covariance matrices. These tests are illustrated in Sec. 7.7. First, one is interested in finding out whether or not the B_{a_i} 's can be pooled. Second, if these matrices can be pooled, one is interested in finding out whether or not the population matrix estimated by the pooled matrix has the required symmetry, i.e., all diagonal elements equal to σ^2 , and all off-diagonal elements equal to $\rho\sigma^2$. In terms of the pooled variance-covariance matrix,

$$MS_{B \times \text{subj w. groups}} = \overline{\text{var}} - \overline{\text{cov}}.$$

For the data under consideration,

$$\overline{\text{var}} - \overline{\text{cov}} = 2.00 - .76 = 1.24.$$

The term $\overline{\text{var}}$ is the mean of entries along the main diagonal of the pooled variance-covariance matrix; $\overline{\text{cov}}$ is the mean of the entries off the main diagonal in this matrix. Also,

$$\overline{\text{var}} + (q - 1) \overline{\text{cov}} = 2.00 + 3(.76) = 4.28.$$

From Table 7.2-4 it is noted that

$$MS_{\text{subj w. groups}} = 4.29,$$

$$MS_{B \times \text{subj w. groups}} = 1.24.$$

The computational formulas for these terms, which are given in Table 7.2-4, are thus short-cut methods. The long method, however, indicates what has to be pooled along the way; the long method also produces summary data that are useful in interpreting the experimental results.

Illustrative Applications. Many applications of this form of the repeated-measures plan will be found in the recent literature in experimental psychology. Shore (1958) reports a study which attempts to evaluate the effects of anxiety level and muscular tension upon perceptual efficiency. Factor *A* in this study was the level of anxiety as measured by the Taylor Manifest Anxiety Scale. On the basis of scores made on this scale, very low (G_1), middle (G_2), and very high (G_3) groups of subjects were formed. There were six subjects in each group. Factor *B* in this study was the level of muscular tension exerted on a dynamometer at the time of the perceptual task. The criterion score was the number of correct symbols recognized in a series of very short exposures. The plan of this experiment may be represented as follows:

Anxiety level	Dynamometer tension					
	b_1	b_2	b_3	b_4	b_5	b_6
a_1	G_1	G_1	G_1	G_1	G_1	G_1
a_2	G_2	G_2	G_2	G_2	G_2	G_2
a_3	G_3	G_3	G_3	G_3	G_3	G_3

Each of the subjects was tested under all levels of dynamometer tension. Level b_6 , the maximum tension condition, was used first for all subjects. The other tension conditions were administered under a restricted randomization procedure. The analysis of the experimental data was reported in a form equivalent to the following (this plan is a special case of a 3×6 factorial experiment with repeated measures on one of the factors, $n = 6$):

	df	MS	<i>F</i>
Between subjects	17		
<i>A</i> (anxiety level)	2	742.7	1.58
Subjects within groups	15	469.0	
Within subjects	90		
<i>B</i> (tension)	5	138.7	3.39**
<i>AB</i>	10	127.6	
<i>B</i> \times subjects within groups	75	40.9	3.12**

The significant interaction indicated that the profiles for the groups had different shapes. Shore interprets this interaction in terms of the group profiles. Note the magnitude of the ratio of the between-subject error term (469.0) to the within-subject error term (40.9). A ratio of 10:1 is not unusual in this area of experimentation.

Another application of this same basic experimental plan is reported by Noble and McNeely (1957). This study was concerned with the relative rate of paired-associates learning as a function of degree of meaningfulness of the pairs. The experimental plan may be represented as follows:

List	Degree of meaningfulness				
	b_1	b_2	b_3	b_4	b_5
a_1	G_1	G_1	G_1	G_1	G_1
a_2	G_2	G_2	G_2	G_2	G_2
\vdots					
a_{18}	G_{18}	G_{18}	G_{18}	G_{18}	G_{18}

Each of the 18 lists had pairs for each of the meaningfulness categories. There were five subjects in each of the groups. Differences between groups of subjects are confounded with differences between the lists; however, primary interest is in the meaningfulness dimension of this experiment. The lists essentially represent samples of material to be learned. The analysis of variance had the following form:

	df	MS	F
Between subjects	89		
<i>A</i> (lists)	17	311.88	1.20
Subjects within groups	72	260.34	
Within subjects	360		
<i>B</i> (meaningfulness)	4	3046.72	52.72**
<i>AB</i>	68	82.91	1.43
<i>B</i> × subjects within groups	288	57.79	

The analysis clearly indicates that degree of meaningfulness of the paired associates is related to the rate of learning.

A study reported by Schrier (1958) provides another illustration of the use of this basic plan. In this study the experimenter was interested in evaluating the effect of amount of reward on the performance of monkeys in a series

of discrimination problems. The criterion used was per cent of correct choices in two periods—a period was a block of 20 consecutive trials. The plan of the experiment may be represented as follows (only one phase of a more extensive experiment is summarized here):

Level of reward	Period	
	b_1	b_2
a_1	G_1	G_1
a_2	G_2	G_2
a_3	G_3	G_3
a_4	G_4	G_4

There were five subjects in each group.

In designs in which there are only two measurements on each subject (and the order in which the measurements are made is not at the option of the experimenter) there is only one covariance. Hence the problem of homogeneity of covariance does not exist. The analysis of the within-subject effects in this case is equivalent to an analysis of difference scores between the two periods.

A scale of measurement in terms of per cent generally does not provide homogeneity of variance. Schrier used an arcsine transformation on the percentages, and the analysis of variance was made on the transformed criterion scale. A modified version of the analysis is given below:

	df	MS	F
Between subjects	19		
A (rewards)	3	148.6	2.52
Subjects within groups	16	58.9	
Within subjects	20		
B (periods)	1	685.8	59.12**
AB	3	1.9	
$B \times$ subjects within groups	16	1.16	

In addition to the analysis indicated above, this worker tested differences in various trends. Tests of the latter kind are discussed in a later section in this chapter.

Another example of the same basic experimental plan is reported by Denenberg and Myers (1958). This study was designed to evaluate the effects of thyroid level upon acquisition and extinction of a bar-pressing

response. Schematically the acquisition phase of this study had the following form:

Thyroid level	Days			
	b_1	b_2	\cdots	b_{15}
a_1	G_1	G_1	\cdots	G_1
a_2	G_2	G_2	\cdots	G_2
a_3	G_3	G_3	\cdots	G_3

There were four subjects in each of the groups. The criterion measure was the number of responses in a Skinner box during a specified time period. The analysis of variance had the following form.

	df	MS	F
Between subjects	11		
A (thyroid level)	2	56,665	13.32**
Subjects within groups	9	4,253	
Within subjects	204		
B (days)	17	594	2.49**
AB	34	1,190	4.98**
$B \times$ subjects within groups	153	239	

Again note the ratio of the between-subject error term (4253) to the within-subject error term (239).

Birch, Burnside, and Clark (1958) report an experiment which had the following form:

Deprivation interval	Blocks of trials			
	b_1	b_2	b_3	
a_1	G_1	G_1	G_1	$n_1 = 10$
a_2	G_2	G_2	G_2	$n_2 = 10$
a_3	G_3	G_3	G_3	$n_3 = 10$
a_4	G_4	G_4	G_4	$n_4 = 9$

There were 10 subjects in each of the groups except G_4 ; there were 9 subjects in the latter groups since 1 subject had to be discarded in the course of the experiment.

Speed of running was used as measure of response strength. The study was designed to investigate response strength as a function of deprivation interval. The analysis of variance had the following form:

	df	MS	F
Between subjects	38		
A (deprivation interval)	3	788.06	3.34*
Subjects within groups	35	236.29	
Within subjects	78		
B (trials)	2	79.92	1.17
AB	6	56.70	
B \times subjects within groups	70	68.24	

In cases of missing data of this kind, where the missing data are the result of conditions unrelated to the treatments, estimates of variation due to treatment effects should be obtained through use of an unweighted means analysis.

General Form of the Expected Values of the Mean Squares. Following the procedures described by Cornfield and Tukey (1956), the expected values for more general cases of the plan under consideration are given below. In the notation used,

$$D_p = 1 - \frac{p}{P}, \quad D_q = 1 - \frac{q}{Q}, \quad D_n = 1 - \frac{n}{N}.$$

Thus each of the D 's is either 0 or 1, depending upon whether the corresponding factor is fixed or random, respectively.

Effect	<i>i</i>	<i>j</i>	<i>k</i>	<i>m</i>	Expected value of mean square
α_i	D_p	q	n	1	$\sigma_\epsilon^2 + D_n D_q \sigma_{\beta\pi}^2 + n D_q \sigma_{\alpha\beta}^2 + q D_n \sigma_\pi^2 + n q \sigma_\alpha^2$
$\pi_{k(i)}$	1	q	D_n	1	$\sigma_\epsilon^2 + D_q \sigma_{\beta\pi}^2 + q D_n \sigma_\pi^2$
β_j	p	D_q	n	1	$\sigma_\epsilon^2 + D_n \sigma_{\beta\pi}^2 + n D_p \sigma_{\alpha\beta}^2 + n p \sigma_\beta^2$
$\alpha\beta_{ij}$	D_p	D_q	n	1	$\sigma_\epsilon^2 + D_n \sigma_{\beta\pi}^2 + n \sigma_{\alpha\beta}^2$
$\beta\pi_{jk(i)}$	1	D_q	D_n	1	$\sigma_\epsilon^2 + D_n \sigma_{\beta\pi}^2$
$\epsilon_{m(ijk)}$	1	1	1	1	σ_ϵ^2

The expected values for the mean squares for the case in which factors A and B are fixed and the subject factor random are obtained by setting $D_p = 0$, $D_q = 0$, and $D_n = 1$.

7.3 Three-factor Experiment with Repeated Measures (Case I)

Two special cases will be considered. The first case will be that of a $p \times q \times r$ factorial experiment in which there are repeated observations on the last two factors. In the second case, repeated measures will be restricted to the last factor. A schematic representation of the first case is given below:

	b_1			\dots	b_q		
	c_1	\dots	c_r		c_1	\dots	c_r
a_1	G_1	\dots	G_1	\dots	G_1	\dots	G_1
a_2	G_2	\dots	G_2	\dots	G_2	\dots	G_2
\vdots							
\vdots							
\vdots							
a_p	G_p	\dots	G_p	\dots	G_p	\dots	G_p

There are n subjects in each group. Each of the subjects is observed under all qr combinations of factor B and C but only under a single level of factor A . Thus there are p groups of n subjects each (np subjects in all); there are qr observations on each subject.

The observations on subjects in group i may be represented as follows:

	Subject	b_1			\dots	b_q			Total
		c_1	\dots	c_r		c_1	\dots	c_r	
a_i	$1(i)$	X_{i111}	\dots	X_{i1r1}	\dots	X_{iq11}	\dots	X_{iqr1}	$P_{1(i)}$
	\vdots								
	$m(i)$	X_{i11m}	\dots	X_{i1rm}	\dots	X_{iq1m}	\dots	X_{iqrm}	$P_{m(i)}$
	\vdots								
	$n(i)$	X_{i11n}	\dots	X_{i1rn}	\dots	X_{iq1n}	\dots	X_{iqrn}	$P_{n(i)}$

The notation X_{iqrm} indicates an observation on subject $m(i)$ under treatment combination abc_{iqr} . The notation $P_{m(i)}$ denotes the sum of the qr observations on subject m in group i . Unless there is some ambiguity about which group is under discussion, the notation P_m will be used to denote this total.

Assuming that A , B , and C are fixed factors, the analysis of variance generally takes the form shown in Table 7.3-1. The expected values of the mean squares indicate appropriate F ratios, provided that the sampling

Table 7.3-1 Summary of Analysis of Variance

Source of variation	df	E(MS) [†]
Between subjects	$np - 1$	
A	$p - 1$	$\sigma_\epsilon^2 + q\sigma_{\alpha\pi}^2 + nq\sigma_{\alpha}^2$
Subj w. groups	$p(n - 1)$	$\sigma_\epsilon^2 + q\sigma_{\pi}^2$
Within subjects	$np(qr - 1)$	
B	$q - 1$	$\sigma_\epsilon^2 + r\sigma_{\beta\pi}^2 + np\sigma_{\beta}^2$
AB	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + r\sigma_{\beta\pi}^2 + nr\sigma_{\alpha\beta}^2$
$B \times$ subj w. groups	$p(n - 1)(q - 1)$	$\sigma_\epsilon^2 + r\sigma_{\beta\pi}^2$
C	$r - 1$	$\sigma_\epsilon^2 + q\sigma_{\gamma\pi}^2 + npq\sigma_{\gamma}^2$
AC	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + q\sigma_{\gamma\pi}^2 + nq\sigma_{\alpha\gamma}^2$
$C \times$ subj w. groups	$p(n - 1)(r - 1)$	$\sigma_\epsilon^2 + q\sigma_{\gamma\pi}^2$
BC	$(q - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\beta\gamma\pi}^2 + np\sigma_{\beta\gamma}^2$
ABC	$(p - 1)(q - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\beta\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$
$BC \times$ subj w. groups	$p(n - 1)(q - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\beta\gamma\pi}^2$

[†] Assumes A , B , and C fixed factors.

distributions of the statistics involved are actually what they are assumed to be. What is assumed will be made explicit later in this section.

An alternative partition of the total degrees of freedom is as follows:

Source of variation	Degrees of freedom	
	General case	Special case $p = 2, q = 3, r = 4, n = 5$
Between cells	$pqr - 1$	23
Within cells	$pqr(n - 1)$	96
Subj w. groups	$p(n - 1)$	8
$B \times$ subj w. groups	$p(n - 1)(q - 1)$	16
$C \times$ subj w. groups	$p(n - 1)(r - 1)$	24
$BC \times$ subj w. groups	$p(n - 1)(q - 1)(r - 1)$	48

A cell in this context contains the n observations under treatment combination abc_{ijk} . There are pqr cells. Treatment and interaction variation are obtained by partitioning the between-cell variation. The latter partition is identical for all $p \times q \times r$ factorial experiments whether or not there

are repeated measures. The manner in which the within-cell variation is partitioned depends upon the pattern of the repeated measures. For the case being considered, each of the parts of the within-cell variation forms a denominator for some F ratio. Hence the following alternative notation system:

$$MS_{\text{subj w. groups}} = MS_{\text{error}(a)},$$

$$MS_{B \times \text{subj w. groups}} = MS_{\text{error}(b)},$$

$$MS_{C \times \text{subj w. groups}} = MS_{\text{error}(c)},$$

$$MS_{BC \times \text{subj w. groups}} = MS_{\text{error}(bc)}.$$

The alternative notation system has the advantage of being more compact as well as more indicative of the role of the error term in the F tests, for the special case being considered. However, the F tests will change as a function of what model is appropriate for the data being analyzed. Thus the alternative notation has the disadvantage of not indicating how each of the terms is computed.

Each of the parts of the within-cell variation may be checked for homogeneity, the latter being one of the assumptions made when the F ratio is considered to be distributed in the form of the F distribution. For example,

Subj w. groups	$p(n-1)$
Subj w. G_1	$n-1$
.....
Subj w. G_p	$n-1$

An F_{\max} test for the homogeneity of the parts would have the critical value $F_{\max(1-\alpha)}(p, n-1)$. As another example,

$BC \times \text{subj w. groups}$	$p(n-1)(q-1)(r-1)$
$BC \times \text{subj w. } G_1$	$(n-1)(q-1)(r-1)$
.....
$BC \times \text{subj w. } G_p$	$(n-1)(q-1)(r-1)$

An F_{\max} test in this last case is, in part, a check on whether or not mean covariances of a set of variance-covariance matrices are equal. This test is only a partial check on whether or not sets of corresponding variance-covariance matrices may be pooled. More appropriate tests are generalizations of those given in Sec. 7.7.

Each of the interactions with subjects may be shown to have the general form

$$\overline{\text{var}} - \overline{\text{cov}}_i$$

with respect to a specified variance-covariance matrix. For example, corresponding to the $BC \times \text{subjects within groups}$ interaction there is a pooled $pq \times pq$ variance-covariance matrix. This matrix is assumed to be

an estimate of a population variance-covariance matrix in which the elements along the main diagonal are all equal to σ^2 and the elements off the main diagonal are all equal to $\rho_i\sigma^2$. For each of the variance-covariance matrices, σ^2 is assumed to be constant, but ρ may vary for the different matrices; that is, ρ is constant within any one matrix but varies for different matrices. If ρ were constant for all matrices, then $\sigma_{\beta\pi}^2 = \sigma_{\gamma\pi}^2 = \sigma_{\beta\gamma\pi}^2$ and all interactions with subjects could be pooled.

If the complete model is appropriate and if there is evidence to indicate that the assumptions with respect to the covariances do not hold, the critical values listed under the heading Conservative Test are more realistic than those given under the heading of Usual Test.

F ratio	Critical values	
	Usual test	Conservative test
$\frac{MS_b}{MS_{B \times \text{subj w. group}}}$	$F_{1-\alpha}[(p-1), p(n-1)(q-1)]$	$F_{1-\alpha}[1, p(n-1)]$
$\frac{MS_{ab}}{MS_{B \times \text{subj w. group}}}$	$F_{1-\alpha}[(p-1)(q-1), p(n-1)(q-1)]$	$F_{1-\alpha}[(p-1), p(n-1)]$
$\frac{MS_c}{MS_{C \times \text{subj w. group}}}$	$F_{1-\alpha}[(r-1), p(n-1)(r-1)]$	$F_{1-\alpha}[1, p(n-1)]$
$\frac{MS_{ac}}{MS_{C \times \text{subj w. group}}}$	$F_{1-\alpha}[(p-1)(r-1), p(n-1)(r-1)]$	$F_{1-\alpha}[(p-1), p(n-1)]$
$\frac{MS_{bc}}{MS_{B \times \text{subj w. group}}}$	$F_{1-\alpha}[(q-1)(r-1), p(n-1)(q-1)(r-1)]$	$F_{1-\alpha}[1, p(n-1)]$
$\frac{MS_{abc}}{MS_{B \times \text{subj w. group}}}$	$F_{1-\alpha}[(p-1)(q-1)(r-1), p(n-1)(q-1)(r-1)]$	$F_{1-\alpha}[(p-1), p(n-1)]$

The expected values for mean squares given in Table 7.3-1 are obtained from a model which includes interactions with the subject factor. If in fact such interactions do not exist (or are negligible relative to the magnitude of σ_e^2), then

$$E(MS_{\text{error}(b)}) = \sigma_e^2 \quad \text{if } \sigma_{\beta\pi}^2 = 0,$$

$$E(MS_{\text{error}(c)}) = \sigma_e^2 \quad \text{if } \sigma_{\gamma\pi}^2 = 0,$$

$$E(MS_{\text{error}(bc)}) = \sigma_e^2 \quad \text{if } \sigma_{\beta\gamma\pi}^2 = 0.$$

In words, if all interactions with the subject factor are zero, each of the above mean squares is an estimate of the same variance, namely, that due to experimental error. Further, these estimates are independent and may be pooled to provide a single estimate of σ_e^2 . Thus,

$$MS_{\text{error(within)}} = \frac{SS_{\text{error}(b)} + SS_{\text{error}(c)} + SS_{\text{error}(bc)}}{p(n-1)(qr-1)}$$

provides an estimate of σ_e^2 having $p(n-1)(qr-1)$ degrees of freedom.

If the experiment provides a relatively large number of degrees of freedom (say, over 30) for estimating the variance due to each of the interactions with subjects, there is generally no need to consider pooling procedures. When there are relatively few degrees of freedom for such estimates, the decision about pooling should depend largely on previous experimental work. In the absence of such background information, preliminary tests on the model are useful. The purpose of such tests is to provide the experimenter with a posteriori information about whether or not certain of the interactions with random factors should be included in the model for the experiment. Such

Table 7.3-2 Denominator of F Ratio for Simple Effects

	Simple effect	Denominator of F ratio
A at b_j	$\overline{AB}_{1j} - \overline{AB}_{2j}$	$[MS_{\text{error}(a)} + (q - 1)MS_{\text{error}(b)}]/q$
A at c_k	$\overline{AC}_{1k} - \overline{AC}_{2k}$	$[MS_{\text{error}(a)} + (r - 1)MS_{\text{error}(c)}]/r$
B at a_i	$\overline{AB}_{i1} - \overline{AB}_{i2}$	$MS_{\text{error}(b)}$
C at a_i	$\overline{AC}_{i1} - \overline{AC}_{i2}$	$MS_{\text{error}(c)}$
B at c_k	$\overline{BC}_{1k} - \overline{BC}_{2k}$	$[MS_{\text{error}(b)} + (r - 1)MS_{\text{error}(bc)}]/r$
C at b_j	$\overline{BC}_{j1} - \overline{BC}_{j2}$	$[MS_{\text{error}(c)} + (q - 1)MS_{\text{error}(bc)}]/q$
A at bc_{jk}	$\overline{ABC}_{1jk} - \overline{ABC}_{2jk}$	$MS_{\text{w, cell}}$

tests should be made at numerically high levels of significance (that is, $\alpha = .20$ or $\alpha = .30$). This procedure does not drop a term from the model unless the data clearly indicate that such terms can be dropped. Since terms are dropped when tests on the model do not reject the hypothesis being tested, high power is required.

Bartlett's test for homogeneity of variance may be used to indicate whether or not interactions with subjects can be pooled. (Pooling is equivalent to dropping terms from the model.) A rough but conservative test for pooling of this kind is to take the ratio of the largest to the smallest of the mean squares due to interactions with subjects. Use as a critical value for this ratio $F_{.90}(df_1, df_2)$, where the df_1 and df_2 are the respective degrees of freedom for the largest and smallest of the respective mean squares. For this procedure, the level of significance is larger than $\alpha = .20$. This test is biased in the direction of rejecting the hypothesis of homogeneity (i.e., does not lead to pooling) more often than an α -level test. The critical value errs on the side of being too small. If the hypothesis of homogeneity is rejected by the rough procedure, Bartlett's test should be used.

Tests on simple main effects have denominators of the form shown in Table 7.3-2. With the exception of simple effects at level a_i , these denominators are actually the pooled error variance for the over-all main effect and the corresponding interaction. For example, the variation due to the

simple main effect of factor A at level b_j is a mixture of the variation due to the main effect of factor A and the variation due to the AB interaction. It is readily shown that

$$\frac{SS_{\text{error}(a)} + SS_{\text{error}(b)}}{p(n-1) + p(n-1)(q-1)} = \frac{MS_{\text{error}(a)} + (q-1)MS_{\text{error}(b)}}{q}.$$

In practice, it is simpler to add the sum of squares and divide by the pooled degrees of freedom than it is to take a weighted mean of the mean squares.

Tests on the difference between simple main effects take the following form:

$$t = \frac{\overline{AB}_{2j} - \overline{AB}_{1j}}{\sqrt{2[MS_{\text{error}(a)} + (q-1)MS_{\text{error}(b)}]/nrq}}$$

$$t = \frac{\overline{AC}_{2k} - \overline{AC}_{1k}}{\sqrt{2[MS_{\text{error}(a)} + (r-1)MS_{\text{error}(c)}]/nrq}}.$$

An approximate critical value for a t statistic of this kind is obtained as follows: Let t_a and t_b be the critical values for a test of level of significance equal to α for the degrees of freedom corresponding to $MS_{\text{error}(a)}$ and $MS_{\text{error}(b)}$, respectively. Then an approximate critical value for the t statistic is

$$t_{\text{critical}} = \frac{t_a MS_{\text{error}(a)} + t_b (q-1) MS_{\text{error}(b)}}{MS_{\text{error}(a)} + (q-1) MS_{\text{error}(b)}}.$$

This critical value is suggested by Cochran and Cox (1957, p. 299). In cases in which the degrees of freedom for the mean squares are both large (say, over 30), the critical value may be obtained directly from tables of the normal distribution.

Computational Procedures. With the exception of the breakdown of the within-cell variation, computational procedures are identical with those of a $p \times q \times r$ factorial experiment having n observations per cell. These procedures will be illustrated by the data in Table 7.3-3.

Table 7.3-3 Basic Data for Numerical Example

Table 1: Basic Data for Numerical Example												
	Subjects	Periods: b_1			b_2			b_3			Total	
		Dials:	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2		c_3
a_1	1		45	53	60	40	52	57	28	37	46	418
	2		35	41	50	30	37	47	25	32	41	338
	3		60	65	75	58	54	70	40	47	50	519
a_2	4		50	48	61	25	34	51	16	23	35	343
	5		42	45	55	30	37	43	22	27	37	338
	6		56	60	77	40	39	57	31	29	46	435

Table 7.3-4 Summary Tables for Numerical Example

ABC summary table

	b_1			b_2			b_3		
	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
a_1	140	159	185	128	143	174	93	116	137
a_2	148	153	193	95	110	151	69	79	118
	288	312	278	223	253	325	162	195	255

AB summary table

	b_1	b_2	b_3	
a_1	484	445	346	1275
a_2	494	356	266	1116
	978	801	612	2391

AC summary table

	c_1	c_2	c_3	
a_1	361	418	496	1275
a_2	312	342	462	1116
	673	760	958	2391

(i)

BC summary table

	c_1	c_2	c_3	
b_1	288	312	378	978
b_2	223	253	325	801
b_3	162	195	255	612
	673	760	958	2391

 $B \times \text{subj w. } G_1$ summary table

Subject	b_1	b_2	b_3	
1	158	149	111	418
2	126	114	98	338
3	200	182	137	519
	484	445	346	1275

 $B \times \text{subj w. } G_2$ summary table

Subject	b_1	b_2	b_3	
4	159	110	74	343
5	142	110	86	338
6	193	136	106	435
	494	356	266	1116

(ii) $C \times \text{subj w. } G_1$ summary table

Subject	c_1	c_2	c_3	
1	113	142	163	418
2	90	110	138	338
3	158	166	195	519
	361	418	496	1275

 $C \times \text{subj w. } G_2$ summary table

Subject	c_1	c_2	c_3	
4	91	105	147	343
5	94	109	135	338
6	127	128	180	435
	312	342	462	1116

Suppose that the levels of factor A represent the noise background under which subjects monitor three dials. The latter define factor C . Subjects are required to make adjustments on the respective dials whenever needles swing outside a specified range. Accuracy scores are obtained for each dial during three consecutive 10-min time periods (factor B). The basic data for this experiment are given in Table 7.3-3. Subjects 1, 2, and 3 make up G_1 ; subjects 4, 5, and 6 make up G_2 . To illustrate the meaning of the data, during the first 10-min interval (b_1) subject 1 had scores of 45, 53, and 60 on dials c_1 , c_2 , and c_3 , respectively.

Summary tables prepared from these basic data are given in Table 7.3-4. In part i are summary tables that would be obtained for any $2 \times 3 \times 3$ factorial experiment having n observations per cell. Part ii is unique to a factorial experiment having repeated measures on factors B and C . In the $B \times$ subjects within G_1 summary table a cell entry will be denoted by the symbol BP_{jm} . For example, $BP_{11} = 158$, and $BP_{13} = 111$. Similarly an entry in a cell of the $C \times$ subjects within group summary table will be denoted by the symbol CP_{km} .

Convenient computational symbols are defined and computed in Table 7.3-5. Symbols (1) through (9) are identical to those used in any $p \times q \times r$ factorial experiment in which there are n observations in each cell. Symbols (10) through (12) are unique to a factorial experiment in which there are repeated measures on factors B and C . By using these symbols, computational formulas take the form given in Table 7.3-6. In terms of means,

$$SS_{B \times \text{subj w. groups}} = r \sum_i \sum_j \sum_k (\overline{BP}_{jm(i)} - \overline{AB}_{im} - \overline{P}_{m(i)} + \overline{A}_i)^2.$$

The formula for this source of variation given in Table 7.3-6 leads to simpler computations. This source of variation is also designated $SS_{\text{error}(b)}$.

There is an over-all computational check that can be made on the sum of the error terms. The following relationship exists:

$$\begin{aligned} SS_{\text{error}(a)} &= (10) - (3) \\ SS_{\text{error}(b)} &= (11) - (6) - (10) + (3) \\ SS_{\text{error}(c)} &= (12) - (7) - (10) + (3) \\ SS_{\text{error}(bc)} &= (2) - (9) - (11) - (12) + (6) + (7) \\ &\quad + (10) - (3) \\ \hline SS_{\text{w. cell}} &= (2) - (9) \end{aligned}$$

The computational symbols on the right may be treated as algebraic symbols; that is, $(3) + (3) = 2(3)$, $(3) - (3) = 0$. The algebraic sum of the symbols on the right is $(2) - (9)$. The latter is the computational formula for $SS_{\text{w. cell}}$. The computational symbols at the bottom of

Table 7.3-5 are used to partition the error variation for each of the error terms into parts which may be checked for homogeneity by means of an F_{\max} test.

The analysis of variance is summarized in Table 7.3-6. If a model which includes interactions with subjects is appropriate for this experiment, the expected values for the mean squares are those shown in Table 7.3-1 and the

Table 7.3-5 Definitions and Numerical Values of Computational Symbols

$(1) = G^2/npqr$	$= (2391)^2/3(2)(3)(3)$	$= 105,868.17$
$(2) = \Sigma X^2$	$= 45^2 + 53^2 + 60^2 + \cdots + 31^2 + 29^2 + 46^2$	$= 115,793$
$(3) = (\Sigma A_i^2)/nqr$	$= (1275^2 + 1116^2)/3(3)(3)$	$= 106,336.33$
$(4) = (\Sigma B_j^2)/npr$	$= (978^2 + 801^2 + 612^2)/3(2)(3)$	$= 109,590.50$
$(5) = (\Sigma C_k^2)/npq$	$= (673^2 + 760^2 + 958^2)/3(2)(3)$	$= 108,238.50$
$(6) = [\Sigma (AB_{ij})^2]/nr$	$= (484^2 + 445^2 + \cdots + 266^2)/3(3)$	$= 110,391.67$
$(7) = [\Sigma (AC_{ik})^2]/nq$	$= (361^2 + 418^2 + \cdots + 462^2)/3(3)$	$= 108,757.00$
$(8) = [\Sigma (BC_{jk})^2]/np$	$= (288^2 + 312^2 + \cdots + 255^2)/3(6)$	$= 111,971.50$
$(9) = [\Sigma (ABC_{ijk})^2]/n$	$= (140^2 + 159^2 + \cdots + 118^2)/3$	$= 112,834.33$
$(10) = (\Sigma P_m^2)/qr$	$= (418^2 + 338^2 + \cdots + 435^2)/3(3)$	$= 108,827.44$
$(11) = [\Sigma (BP_{jm})^2]/r$	$= (158^2 + 149^2 + \cdots + 106^2)/3$	$= 113,117.67$
$(12) = [\Sigma (CP_{km})^2]/q$	$= (113^2 + 142^2 + \cdots + 180^2)/3$	$= 111,353.67$
<hr/>		
$(10a_1) = \left(\Sigma_{a_1} P_m^2 \right) / qr$	$= (418^2 + 338^2 + 519^2)/3(3)$	$= 62,036.56$
$(10a_2) = \left(\Sigma_{a_2} P_m^2 \right) / qr$	$= (343^2 + 338^2 + 435^2)/3(3)$	$= 46,790.89$
		$108,827.45$
$(11a_1) = \left[\Sigma_{a_1} (BP_{jm}^2) \right] / r$	$= (158^2 + 149^2 + \cdots + 137^2)/3$	$= 63,285.00$
$(11a_2) = \left[\Sigma_{a_2} (BP_{jm}^2) \right] / r$	$= (159^2 + 110^2 + \cdots + 106^2)/3$	$= 49,832.67$
		$113,117.67$

structure of F ratios is determined by these expected values. However, should the existence of interactions with subjects be open to question, preliminary tests on the model are appropriate. In this case such interactions have relatively few degrees of freedom. A check on the homogeneity of such interactions is carried out in Table 7.3-7, by use of Bartlett's test. Since the observed chi square (4.77) exceeds the critical value (3.22) for a test with $\alpha = .20$, the test indicates that the interactions should not be pooled. Equivalently, the test indicates that interactions with subjects should not be dropped from the model.

Table 7.3-6 Summary of Analysis of Variance

Source of variation	Computational formula	SS	df	MS	F
Between subjects		2959.27	5		
<i>A</i>	$\frac{(10) - (1)}{(3) - (1)}$	468.16	1	468.16	
Subj w. groups [error (<i>a</i>)]	$(10) - (3)$	2491.11	4	622.78	
Within subjects		6965.56	48		
<i>B</i>	$\frac{(4) - (1)}{(6) - (3) - (4) + (1)}$	3722.33	2	1861.16	63.39*
<i>AB</i>	$(11) - (6) - (10) + (3)$	333.00	2	166.50	5.67*
<i>B</i> × subj w. groups [error (<i>b</i>)]		234.89	8	29.36	
<i>C</i>	$\frac{(5) - (1)}{(7) - (3) - (5) + (1)}$	2370.33	2	1185.16	89.78*
<i>AC</i>	$(12) - (7) - (10) + (3)$	50.34	2	25.17	1.91
<i>C</i> × subj w. groups [error (<i>c</i>)]		105.56	8	13.20	
<i>BC</i>	$\frac{(8) - (4) - (5) + (1)}{(9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)}$	10.67	4	2.67	
<i>ABC</i>	$(2) - (9) - (11) - (12) + (6) + (7) + (10) - (3)$	11.32	4	2.83	
<i>BC</i> × subj w. groups [error (<i>bc</i>)]		127.11	16	7.94	

Thus tests on B and AB use $MS_{\text{error}(b)}$ as a denominator for F ratios; tests on C and AC use $MS_{\text{error}(c)}$ as a denominator; tests on BC and ABC use $MS_{\text{error}(bc)}$ as a denominator. The main effect of A is tested with $MS_{\text{error}(a)}$. By using $\alpha = .05$ for all tests, the main effects for factors B (periods) and C (dials) are found to be statistically significant. Inspection of the totals for levels b_1 , b_2 , and b_3 indicates decreasing accuracy scores for

Table 7.3-7 Test for Homogeneity of Interactions with Subjects

SS	MS	df	log MS	1/df
234.89	29.36	8	1.468	.125
105.56	13.20	8	1.121	.125
127.11	7.94	16	.900	.062
$\Sigma SS = 467.56$		$\Sigma df = 32$		$\Sigma (1/df) = .312$

$$MS_{\text{pooled}} = (\Sigma SS)/\Sigma df = 467.56/32 = 14.61$$

$$A = \Sigma [(df)_i \log MS_i] = 8(1.468) + 8(1.121) + 16(.900) = 35.112$$

$$B = (\Sigma df) \log MS_{\text{pooled}} = 32(1.165) = 37.280$$

$$C = 1 + \frac{1}{3(k-1)} [\Sigma (1/df) - (1/\Sigma df)] = 1 + \frac{1}{6} [.312 - .031] = 1.047$$

$$\chi^2 = \frac{2.303(B - A)}{C} = \frac{2.303(37.280 - 35.112)}{1.047} = 4.77$$

$$\chi^2_{.80}(2) = 3.22$$

the consecutive time periods. Inspection of the totals for the dials indicates that dial c_3 is monitored with the greatest accuracy and dial c_1 monitored with the least accuracy.

The AB interaction is also noted to be statistically significant. The profiles of means corresponding to the cell totals in the AB summary table are plotted in Fig. 7.1-3. The profiles indicate a difference in the rate of decline in the accuracy scores in the three periods, the group working under noise level a_1 showing a slower decline rate than the group working under a_2 . Differences between corresponding points on these profiles have the form

$$\overline{AB}_{1j} - \overline{AB}_{2j}.$$

The standard error of the difference between these two means is estimated by

$$\begin{aligned} \sqrt{\frac{2(SS_{\text{subj w. groups}} + SS_{B \times \text{subj w. groups}})}{nr[p(n-1) + p(n-1)(q-1)]}} &= \sqrt{\frac{2(2491.11 + 234.89)}{9(12)}} \\ &= \sqrt{50.48} = 7.10. \end{aligned}$$

By way of contrast, the standard error of the difference between two means of the following form,

$$\overline{AB}_{i1} - \overline{AB}_{i2},$$

is estimated by

$$\sqrt{2MS_{\text{error}(b)}/nr} = \sqrt{2(29.36)/9} = 2.55.$$

The latter standard error is considerably smaller than that computed in the last paragraph. A difference of the form $\overline{AB}_{1j} - \overline{AB}_{2j}$ is in part confounded with between-group effects, whereas a difference of the form $\overline{AB}_{i1} - \overline{AB}_{i2}$ is entirely a within-subject effect.

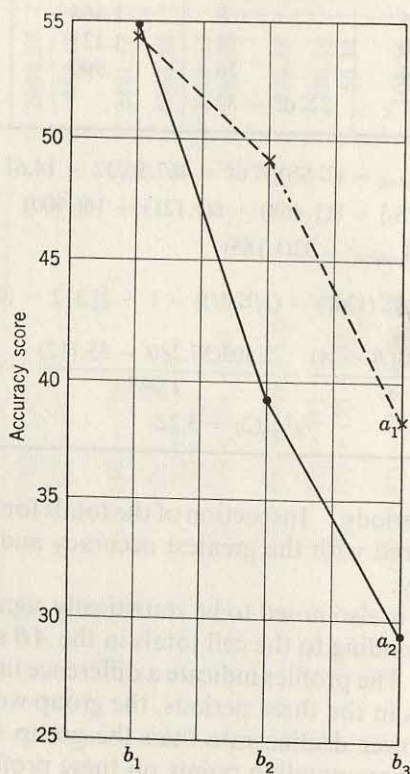


Figure 7.1 Profiles of means.

Illustrative Applications. Many examples of this basic plan will be found in the recent experimental literature. A study by French (1959) illustrates one application of this plan. The purpose of this study was to investigate the effect of lesions in two different areas of the cortex (factor *A*) upon a bar-pressing task which set up two conditions of illumination (factor *B*). Each of the subjects was given three blocks of trials (factor *C*) under each illumination condition. There were four subjects per group. The

criterion measure was the average length of time the bar was depressed during each block of trials. To obtain homogeneity of error variance, the scale of measurement was transformed into $\log(\bar{X} + 1)$. The addition of unity to each of the means avoids the occurrence of the logarithm of zero. The plan for the experiment may be represented as follows:

Lesions	Illumination: b_1			b_2			
	Trials: c_1 c_2 c_3			c_1	c_2	c_3	
a_1	G_1	G_1	G_1	G_1	G_1	G_1	$n = 4$
a_2	G_2	G_2	G_2	G_2	G_2	G_2	

The analysis of variance reported by French had the form given in Table 7.3-8.

In this analysis all interactions with subjects were pooled. The resulting pooled interaction term was the denominator for all within-subject effects. The significant main effect due to factor A indicated that the groups differed

Table 7.3-8 Analysis of Variance for French Data (1959)

Source	df	MS	F
Between subjects	7		
A (lesions)	1	.846	6.89*
Subj w. groups	6	.122	
Within subjects	40		
B (illumination)	1	.161	1.46
C (trials)	2	.945	8.61*
AB	1	.008	
AC	2	.389	3.54*
BC	2	.113	1.03
ABC	2	.078	
Pooled interactions with subjects	30	.110	

in their mean over-all criterion scores (on the transformed scale of measurement). The significant main effect due to trials indicated that the means changed during the trials. The significant lesion \times trial (AC) interaction indicated that the rate of change during the trials differed for the two groups.

Briggs, Fitts, and Bahrick (1958) have reported a series of experiments in which the experimental plan under discussion was used. In one of these studies, the effects of methods of training (factor A) upon transfer of training in a complex tracking task were studied. Performance under the last two blocks of trials in the training session and the first two blocks of the transfer session was analyzed. The plan of the experiment may be represented schematically as follows:

Method of training	Training		Transfer	
	Trial c_1	Trial c_2	Trial c_1	Trial c_2
a_1	G_1	G_1	G_1	G_1
a_2	G_2	G_2	G_2	G_2
a_3	G_3	G_3	G_3	G_3
a_4	G_4	G_4	G_4	G_4

The last two blocks of trials under training and the first two trials under transfer define the two levels of factor B . There were 17 subjects in each group. This plan may be considered as a special case of a $4 \times 2 \times 2$ factorial experiment with repeated measures on the last two factors. The criterion measure was average time on target. The analysis of variance reported by these workers had the form given in Table 7.3-9.

Table 7.3-9 Analysis of Variance for Briggs, Fitts, and Bahrck Data (1958)

Source	df	MS	F
Between subjects	67		
A (methods of training)	3	2,319.02	3.72*
Subj w. groups	64	623.72	
Within subjects	204		
B (training-transfer)	1	11,492.52	113.11*
AB	3	707.44	
$B \times$ subj w. groups	64	101.61	6.96*
C (blocks of trials)	1	5.95	
AC	3	9.20	2.06
$C \times$ subj w. groups	64	90.45	
BC	1	111.90	2.06
ABC	3	24.84	
$BC \times$ subj w. groups	64	54.28	

In the actual experiment, G_1 worked under identical conditions during the training and transfer trial. Each of the other groups had different conditions under the learning trials but a common condition under the transfer trials. The following means were obtained from the experimental data.

	b_1	b_2
a_1	2.36	2.54
a_2	1.80	2.36
a_3	1.45	2.21
a_4	1.54	2.31

Because of the significant interaction, key simple main effects were tested. Means within brackets do not differ significantly, while those means not bracketed do differ at the .05 level. It is noted that all groups were equally good, on the average, on the transfer block of trials. The significant AB interaction was due largely to the fact that the control group (a_1) showed no statistically different increase from training trials (b_1) to transfer trials (b_2), whereas all the other groups did show a statistically significant increase.

Another study reported by Briggs, Fitts, and Bahrck (1958) may be considered as a special case of a $4 \times 3 \times 3$ factorial experiment with repeated measures on the last two factors. In this study, factor A defines the time at which transfer was made from a common training condition to a common transfer condition. Group 4 made the transfer at time zero; i.e., group 4 worked under the transfer condition during all sessions. A schematic representation of the plan is given below.

Transfer time	Sessions: b_1				b_2				b_3			
	Trials: c_1 c_2 c_3 c_4				c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
a_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1
a_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2
a_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3	G_3
a_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4	G_4

$$n_1 = 14, \quad n_2 = 12, \quad n_3 = 14, \quad n_4 = 13$$

Because of unequal group sizes, the equivalent of an unweighted-means analysis was made on the data. In this analysis, group means replace each of the X 's and the data are considered as if n were unity in the calculation of the parts of the between-cell variation. Sums of squares computed in this manner are then multiplied by the harmonic mean of the group sizes.

The subject within groups variation is given by

$$SS_{\text{subj.w. groups}} = \frac{\sum P_m^2}{qr} - \sum \frac{A_i^2}{n_i}$$

The $B \times$ subject within groups interaction is most easily obtained by computing the separate terms $B \times$ subject within G_i and then pooling the parts. Other interactions with subjects are computed in a similar way. The analysis of variance reported by these workers had the form given in Table 7.3-10. Significant interactions led to a series of tests on simple effects as well as individual comparisons between pairs of means. Graphs of profiles corresponding to interactions were used effectively in the interpretations.

A study by Casteneda and Lipsett (1959) provides another illustrative example of the same basic experimental plan. This study was designed to evaluate the effect of level of stress (factor *A*) upon learning patterns of varying complexity (factor *B*). Four block of trials (factor *C*) were given

Table 7.3-10 Analysis of Variance for Briggs, Fitts, and Bahrck Data (1958)

Source	df	MS	<i>F</i>
Between subjects	52		
<i>A</i> (transfer time)	3	45.79	
Subj w. groups	49	69.76	
Within subjects	583		
<i>B</i> (sessions)	2	426.51	17.28*
<i>AB</i>	6	266.12	10.78*
<i>B</i> × subj w. groups	98	24.68	
<i>C</i> (trials)	3	144.11	46.04*
<i>AC</i>	9	2.34	
<i>C</i> × subj w. groups	147	3.13	
<i>BC</i>	6	16.45	5.24*
<i>ABC</i>	18	10.61	3.38*
<i>BC</i> × subj w. groups	294	3.14	

under each level of complexity. The criterion measure was the number of correct responses in each block of trials. The plan may be represented as follows:

Level of stress	Pattern complexity: b_1					b_2			
	Trials:	c_1	c_2	c_3	c_4	c_1	c_2	c_3	c_4
a_1		G_1	G_1	G_1	G_1	G_1	G_1	G_1	G_1
a_2		G_2	G_2	G_2	G_2	G_2	G_2	G_2	G_2

There were 54 subjects in each group. The order of presentation of the patterns was randomized, but each pattern was presented an equal number of times.

The analysis of variance reported by these workers had the form given in Table 7.3-11. It will be noted that relative to the within-subject error terms the between-subject error term is exceptionally large ($MS_{\text{subj w. groups}} = 804.66$). This observation and remarks made by the workers about the distributions suggest that the analysis of variance might more appropriately have been made in terms of a different scale of measurement, that

is, $\log X$ rather than X . In the analysis as it stands, the stress \times pattern interaction (AB) was found to be statistically significant. In terms of the experimental variables, this finding indicated that the stress condition interfered more in the learning of the complex pattern than it did in the learning of the simple pattern. This essentially was the hypothesis that the workers were interested in testing. Because of the extremely high

Table 7.3-11 Analysis of Variance for Casteneda and Lipsett Data (1959)

Source	df	MS	F
Between subjects	107		
<i>A</i> (stress)	1	0.78	
Subj w. groups	106	804.66	
Within subjects	756		
<i>B</i> (patterns)	1	1820.04	255.26*
<i>AB</i>	1	148.34	20.80*
<i>B</i> \times subj w. groups	106	7.13	
<i>C</i> (trials)	3	167.13	80.40*
<i>AC</i>	3	4.01	2.02
<i>C</i> \times subj w. groups	318	1.98	
<i>BC</i>	3	7.38	2.64
<i>ABC</i>	3	6.48	2.32
<i>BC</i> \times subj w. groups	318	2.79	

between-subject error term simple main effects of the stress conditions could not be demonstrated to be statistically significant.

General Expected Values for the Mean Squares. The expected values for the mean squares given in Table 7.3-1 are for the special case in which A , B , and C are fixed factors. The general case is given in Table 7.3-12. In this table,

$$D_p = 1 - \frac{p}{P}, \quad D_q = 1 - \frac{q}{Q}, \quad D_r = 1 - \frac{r}{R}.$$

Thus D_p is either 0 or 1 depending upon whether factor A is fixed or random, respectively.

If $D_p = 1$, $D_q = 0$, and $D_r = 0$, then the special case given in Table 7.3-13 is obtained. There are several assumptions about the form of variance-covariance matrices that must be met in order that these expected values provide reasonable guides for setting up F ratios. A series of preliminary tests on the model may be made, in the absence of background information, in order to check the desirability of pooling various interactions with random factors.

Table 7.3-12 General Expected Values for Mean Squares

Effect	<i>i</i>	<i>j</i>	<i>k</i>	<i>m</i>	<i>o</i>	E(MS)
α_i	D_p	q	r	n	1	$\sigma_\varepsilon^2 + D_q D_r \sigma_{\beta\gamma\pi}^2 + n D_q D_r \sigma_{\alpha\beta\gamma}^2 + q D_r \sigma_{\gamma\pi}^2 + n q D_r \sigma_{\alpha\gamma}^2 + r D_q \sigma_{\beta\pi}^2$
$\pi_{m(i)}$	1	q	r	1	1	$\sigma_\varepsilon^2 + D_q D_r \sigma_{\beta\gamma\pi}^2 + q D_r \sigma_{\gamma\pi}^2 + r D_q \sigma_{\beta\pi}^2 + q r \sigma_\pi^2 + n r D_q \sigma_{\alpha\beta}^2 + n q r \sigma_\alpha^2$
β_j	p	D_a	r	n	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + n D_p D_r \sigma_{\alpha\beta\gamma}^2 + n p D_r \sigma_{\beta\gamma}^2 + r \sigma_{\beta\pi}^2 + n r D_p \sigma_{\alpha\beta}^2 + n p r \sigma_\beta^2$
$\alpha\beta_{ij}$	D_p	D_a	r	n	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + n D_r \sigma_{\alpha\beta\gamma}^2 + r \sigma_{\beta\pi}^2 + n r \sigma_{\alpha\beta}^2$
$\beta\pi_{jm(i)}$	1	D_a	r	1	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + r \sigma_{\beta\pi}^2$
γ_k	p	q	D_r	n	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + n D_p D_q \sigma_{\alpha\beta\gamma}^2 + n p D_q \sigma_{\beta\gamma}^2 + q \sigma_{\gamma\pi}^2 + n q D_p \sigma_{\alpha\gamma}^2 + n p q \sigma_\gamma^2$
$\alpha\gamma_{ik}$	D_p	q	D_r	n	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + n D_q \sigma_{\alpha\beta\gamma}^2 + q \sigma_{\gamma\pi}^2 + n q \sigma_{\alpha\gamma}^2$
$\gamma\pi_{km(i)}$	1	q	D_r	1	1	$\sigma_\varepsilon^2 + D_q \sigma_{\beta\gamma\pi}^2 + q \sigma_{\gamma\pi}^2$
$\beta\gamma_{jk}$	p	D_a	D_r	n	1	$\sigma_\varepsilon^2 + \sigma_{\beta\gamma\pi}^2 + n D_p \sigma_{\alpha\beta\gamma}^2 + n p \sigma_{\beta\gamma}^2$
$\alpha\beta\gamma_{ijk}$	D_p	D_a	D_r	n	1	$\sigma_\varepsilon^2 + \sigma_{\beta\gamma\pi}^2 + n \sigma_{\alpha\beta\gamma}^2$
$\beta\gamma\pi_{jkm(i)}$	1	D_a	D_r	1	1	$\sigma_\varepsilon^2 + \sigma_{\beta\gamma\pi}^2$
$\varepsilon_{o(ijkm)}$	1	1	1	1	1	σ_ε^2

Table 7.3-13 Expected Values of Mean Squares for Case in Which Factor *A* Is Random and Factors *B* and *C* Are Fixed

Source	E(MS)
Between subjects	
<i>A</i>	$\sigma_e^2 + qr\sigma_\pi^2 + nqr\sigma_\alpha^2$
Subj w. groups	$\sigma_e^2 + qr\sigma_\pi^2$
Within subjects	
<i>B</i>	$\sigma_e^2 + r\sigma_{\beta\pi}^2 + nr\sigma_{\alpha\beta}^2 + npr\sigma_\beta^2$
<i>AB</i>	$\sigma_e^2 + r\sigma_{\beta\pi}^2 + nr\sigma_{\alpha\beta}^2$
<i>B</i> × subj w. groups	$\sigma_e^2 + r\sigma_{\beta\pi}^2$
<i>C</i>	$\sigma_e^2 + q\sigma_{\gamma\pi}^2 + nq\sigma_{\alpha\gamma}^2 + npq\sigma_\gamma^2$
<i>AC</i>	$\sigma_e^2 + q\sigma_{\gamma\pi}^2 + nq\sigma_{\alpha\gamma}^2$
<i>C</i> × subj w. groups	$\sigma_e^2 + q\sigma_{\gamma\pi}^2$
<i>BC</i>	$\sigma_e^2 + \sigma_{\beta\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2 + np\sigma_{\beta\gamma}^2$
<i>ABC</i>	$\sigma_e^2 + \sigma_{\beta\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$
<i>BC</i> × subj w. groups	$\sigma_e^2 + \sigma_{\beta\gamma\pi}^2$

7.4 Three-factor Experiment with Repeated Measures (Case II)

In the last section the case in which there were repeated measures on two of the three factors was considered. In this section the case in which there are repeated measures on only one of the three factors will be considered. This case may be represented schematically as follows:

		c_1	c_2	\cdots	c_r
a_1	b_1	G_{11}	G_{11}	\cdots	G_{11}
	\vdots				
	\vdots				
	b_q	G_{1q}	G_{1q}	\cdots	G_{1q}
a_p	\vdots				
	\vdots				
	b_1	G_{p1}	G_{p1}	\cdots	G_{p1}
	\vdots				
	b_q	G_{pq}	G_{pq}	\cdots	G_{pq}

Each of the groups is observed under all levels of factor *C*, but each group is assigned to only one combination of factors *A* and *B*. The notation G_{ij} denotes the group of subjects assigned to treatment combination ab_{ij} . A

subject within group G_{ij} is identified by the subscript $m(ij)$. This notation indicates that the subject effect is nested under both factors A and B .

The structural model on which the analysis which follows is based has the following form:

$$E(X_{ijkm}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \pi_{m(ij)} \\ + \gamma_k + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \gamma\pi_{km(ij)}.$$

Since the subject factor is nested under both factors A and B , there can be no interaction between these latter factors and the subject factor. This model has implicit in it homogeneity assumptions on variance-covariance matrices associated with the repeated measures. The analysis of variance for this plan takes the form given in Table 7.4-1. The expected values in this table are for the special case in which A , B , and C are considered fixed factors.

Table 7.4-1 Summary of Analysis of Variance

Source of variation	df	E(MS)†
Between subjects	$npq - 1$	
A	$p - 1$	$\sigma_\epsilon^2 + r\sigma_\pi^2 + nq\sigma_\alpha^2$
B	$q - 1$	$\sigma_\epsilon^2 + r\sigma_\pi^2 + np\sigma_\beta^2$
AB	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + r\sigma_\pi^2 + nr\sigma_{\alpha\beta}^2$
Subj w. groups [error (between)]	$pq(n - 1)$	$\sigma_\epsilon^2 + r\sigma_\pi^2$
Within subjects	$npq(r - 1)$	
C	$r - 1$	$\sigma_\epsilon^2 + \sigma_{\gamma\pi}^2 + npq\sigma_\gamma^2$
AC	$(p - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\gamma\pi}^2 + nq\sigma_{\alpha\gamma}^2$
BC	$(q - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\gamma\pi}^2 + np\sigma_{\beta\gamma}^2$
ABC	$(p - 1)(q - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$
$C \times$ subj w. groups [error (within)]	$pq(n - 1)(r - 1)$	$\sigma_\epsilon^2 + \sigma_{\gamma\pi}^2$

† Assumes A , B , and C fixed factors.

An alternative partition of the total variation permits a comparison between this plan and a $p \times q \times r$ factorial experiment in which there are no repeated measures, but n observations per cell.

Source	df
Total	$npqr - 1$
Between cells	$pqr - 1$
Within cells	$pqr(n - 1)$
Subj w. groups	$pq(n - 1)$
$C \times$ subj w. groups	$pq(n - 1)(r - 1)$

The main effects and all interactions of factors A , B , and C are part of the between-cell variation whether or not there are repeated measures. The partition of the between-cell variation is identical in the two cases. When there are repeated measures on factor C , the within-cell variation is subdivided into two parts. One of these parts is

$$\begin{aligned} SS_{\text{subj w. groups}} &= r \sum \sum \sum (\bar{P}_{m(ij)} - \overline{AB}_{ij})^2 \\ &= \frac{\sum \sum \sum (P_{m(ij)}^2)}{r} - \frac{\sum \sum (AB_{ij})^2}{nr}. \end{aligned}$$

(The symbol $\sum \sum \sum P_{m(ij)}^2$ represents the sum of the squared totals from each subject. Each total is based upon r observations.) This source of variation is a measure of the extent to which the mean of a subject differs from the mean of the group in which the subject is located. The other part of the within-cell variation is

$$SS_{C \times \text{subj w. groups}} = SS_{\text{w. cell}} - SS_{\text{subj w. groups}},$$

where

$$\begin{aligned} SS_{\text{w. cell}} &= \sum \sum \sum \sum (X_{ijkm} - \overline{ABC}_{ijk})^2 \\ &= \sum \sum \sum \sum X_{ijkm}^2 - \frac{\sum (ABC_{ijk}^2)}{n}. \end{aligned}$$

Because of the structure of the F ratio for this plan (when A , B , and C are fixed factors), the following notation is sometimes used:

$$SS_{\text{subj w. groups}} = SS_{\text{error(between)}},$$

$$SS_{C \times \text{subj w. groups}} = SS_{\text{error(within)}}.$$

Each of these error terms may be subdivided and tested for homogeneity by means of F_{\max} tests. The first term may be subdivided into the following parts:

Source	df
Subj w. G_{11}	$n - 1$
Subj w. G_{12}	$n - 1$
.....
Subj w. G_{pq}	$n - 1$

There will be pq terms, each having the general form

$$SS_{\text{subj w. } G_{ij}} = r \sum_m (\bar{P}_{m(ij)} - \overline{AB}_{ij})^2.$$

The critical value for an F_{\max} test would be $F_{\max(1-\alpha)}(pq, n-1)$. The second error term may be divided into the following parts:

Source	df
$C \times \text{subj w. } G_{11}$	$(n-1)(r-1)$
$C \times \text{subj w. } G_{12}$	$(n-1)(r-1)$
.....
$C \times \text{subj w. } G_{pq}$	$(n-1)(r-1)$

There will be pq terms; each has the general form

$$SS_{C \times \text{subj w. } G_{ij}} = \sum_m \sum_k (X_{ijkm} - \bar{P}_{m(ij)} - \bar{ABC}_{ijk} + \bar{AB}_{ij})^2.$$

The critical value for an F_{\max} test in this case would be

$$F_{\max(1-\alpha)}[pq, (n-1)(r-1)].$$

Should either of these two error terms prove to be heterogeneous in terms of the criterion scale of measurement being used, the experimenter should consider a transformation on the scale of measurement in terms of which the analysis of variance may be carried out.

Table 7.4-2 Denominator of F Ratio for Simple Effects

Simple effect	Denominator of F ratio
$A \text{ at } b_j \quad \bar{AB}_{1j} - \bar{AB}_{2j}$	} $MS_{\text{error(between)}}$
$B \text{ at } a_i \quad \bar{AB}_{i1} - \bar{AB}_{i2}$	
$C \text{ at } a_i \quad \bar{AC}_{i1} - \bar{AC}_{i2}$	} $MS_{\text{error(within)}}$
$C \text{ at } b_j \quad \bar{BC}_{j1} - \bar{BC}_{j2}$	
$C \text{ at } ab_{ij} \quad \bar{ABC}_{ij1} - \bar{ABC}_{ij2}$	
$A \text{ at } c_k \quad \bar{AC}_{1k} - \bar{AC}_{2k}$	} $[MS_{\text{error(between)}} + (r-1)MS_{\text{error(within)}}]/r$
$B \text{ at } c_k \quad \bar{BC}_{1k} - \bar{BC}_{2k}$	
$AB \text{ at } c_k \quad \bar{ABC}_{12k} - \bar{ABC}_{34k}$	

In making tests on simple main effects, denominators appropriate for F ratios are indicated in Table 7.4-2. It should be noted that a difference between two simple main effects is a mixture of main effects and interaction effects. In cases where the main effects and interaction effects have different error terms, a compromise error term is constructed. The latter is a weighted average of the different error terms, the weights being the respective degrees of freedom. Because of this pooling of heterogeneous sources of variation, the resulting F tests are potentially subject to bias.

In this design, when the assumptions of homogeneity of covariances are questionable, critical values of the conservative tests involving factor C have the form

$$F_{1-\alpha}[1, pq(n-1)] \quad \text{instead of} \quad F_{1-\alpha}[(r-1), pq(n-1)(r-1)],$$

$$F_{1-\alpha}[(p-1), pq(n-1)]$$

$$\text{instead of} \quad F_{1-\alpha}[(p-1)(r-1), pq(n-1)(r-1)].$$

That is, the degrees of freedom for numerator and denominator are each divided by $r-1$.

Computational Procedures. A numerical example will be used to illustrate the computational procedures. [This example is a modified version of an experiment actually conducted by Meyer and Noble (1958). Suppose that an experimenter is interested in evaluating the effect of anxiety (factor A) and muscular tension (factor B) on a learning task. Subjects who score extremely low on a scale measuring manifest anxiety are assigned to level a_1 ; subjects who score extremely high are assigned to level a_2 . The

Table 7.4-3 Basic Data for Numerical Example

		Subjects	Blocks of trials				Total
			c_1	c_2	c_3	c_4	
a_1	b_1	1	18	14	12	6	50
		2	19	12	8	4	43
		3	14	10	6	2	32
	b_2	4	16	12	10	4	42
		5	12	8	6	2	28
		6	18	10	5	1	34
a_2	b_1	7	16	10	8	4	38
		8	18	8	4	1	31
		9	16	12	6	2	36
	b_2	10	19	16	10	8	53
		11	16	14	10	9	49
		12	16	12	8	8	44

tension factor is defined by pressure exerted on a dynamometer. One half of the subjects at level a_1 are assigned at random to tension condition b_1 ; the other half are assigned to level b_2 . The subjects at level a_2 are divided in a similar manner. Subjects are given four blocks of trials (factor C). The criterion is the number of errors in each block of trials. Suppose that the observed data are those given in Table 7.4-3.

In this table subjects 1, 2, and 3 form group G_{11} ; subjects 4, 5, and 6 form group G_{12} ; etc. Subject 6 is represented symbolically as $P_{3(12)}$, that is, the third subject in G_{12} . This plan may be classified as a $2 \times 2 \times 4$ factorial experiment with repeated measures on the last factor, $n = 3$. Summary data obtained from the basic observations are given in Table 7.4-4. All these summary tables are identical to those that would be obtained for a $2 \times 2 \times 4$ factorial experiment having no repeated measures. Computational symbols are defined and evaluated in Table 7.4-5. Symbol (10) is the only one unique to a repeated-measures design. The data for the latter symbol are obtained from the Total column in Table 7.4-3.

The analysis of variance is summarized in Table 7.4-6. Suppose that the

Table 7.4-4 Summary Table for Numerical Example

ABC summary table

		c_1	c_2	c_3	c_4	Total
a_1	b_1	51	36	26	12	125
	b_2	46	30	21	7	104
a_2	b_1	50	30	18	7	105
	b_2	51	42	28	25	146
		198	138	93	51	480

AB summary table

	b_1	b_2	Total
a_1	125	104	229
a_2	105	146	251
	230	250	480

AC summary table

	c_1	c_2	c_3	c_4	Total
a_1	97	66	47	19	229
a_2	101	72	46	32	251
	198	138	93	51	480

BC summary table

	c_1	c_2	c_3	c_4	Total
b_1	101	66	44	19	230
b_2	97	72	49	32	250
	198	138	93	51	480

Table 7.4-5 Definitions and Numerical Values of Computational Symbols

(1) = $G^2/npqr$	= $(480)^2/48$	= 4800.00
(2) = ΣX^2	= $18^2 + 14^2 + \dots + 8^2 + 8^2$	= 6058
(3) = $(\Sigma A_i^2)/nqr$	= $(229^2 + 251^2)/24$	= 4810.08
(4) = $(\Sigma B_j^2)/npq$	= $(230^2 + 250^2)/24$	= 4808.33
(5) = $(\Sigma C_k^2)/npq$	= $(198^2 + 138^2 + 93^2 + 51^2)/12$	= 5791.50
(6) = $[\Sigma(AB_{ij}^2)]/nr$	= $(125^2 + 104^2 + 105^2 + 146^2)/12$	= 4898.50
(7) = $[\Sigma(AC_{ik}^2)]/nq$	= $(97^2 + 66^2 + \dots + 46^2 + 32^2)/6$	= 5810.00
(8) = $[\Sigma(BC_{jk}^2)]/np$	= $(101^2 + 66^2 + \dots + 49^2 + 32^2)/6$	= 5812.00
(9) = $[\Sigma(ABC_{ijk}^2)]/n$	= $(51^2 + 36^2 + \dots + 28^2 + 25^2)/3$	= 5923.33
(10) = $(\Sigma P_m^2)/r$	= $(50^2 + 43^2 + \dots + 49^2 + 44^2)/4$	= 4981.00

Table 7.4-6 Summary of Analysis of Variance

Source of variation	Computational formula	SS	df	MS	F
Between subjects		181.00	11		
A (anxiety)	$(10) - (1)$			10.08	
B (tension)	$(3) - (1)$	10.08	1	8.33	
AB	$(4) - (1)$	8.33	1	80.09	7.77*
Subj w. groups [error (between)]	$(6) - (3) - (4) + (1)$	80.09	1		
	$(10) - (6)$	82.50	8	10.31	
Within subjects		1077.00	36		
C (trials)	$(2) - (10)$	991.50	3	330.50	152.30*
AC	$(5) - (1)$	8.42	3	2.81	1.29
BC	$(7) - (3) - (5) + (1)$	12.17	3	4.06	1.87
ABC	$(8) - (4) - (5) + (1)$	12.74	3	4.25	1.96
C \times subj w. groups [error (within)]	$(9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)$				
	$(2) - (9) - (10) + (6)$	52.17	24	2.17	

.05 level of significance is used in all tests. The main effect for factor *C* (trials) is found to be statistically significant. This indicates that the average number of errors differed in the four blocks of trials. Inspection of the totals for the blocks indicates a decreasing number of errors from c_1 to c_4 . The anxiety \times tension interaction is also statistically significant. This indicates that the pattern of the number of errors in the two anxiety groups

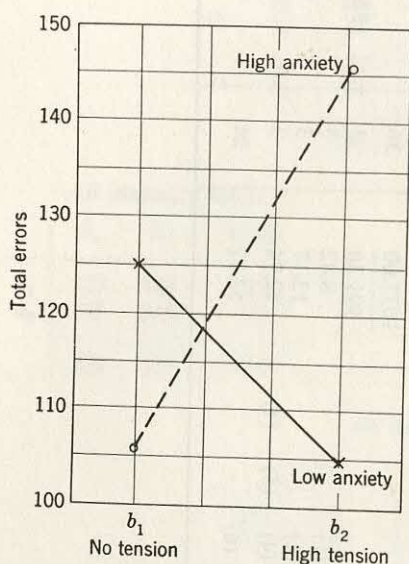


Figure 7.2

depends upon the level of muscular tension. The profiles corresponding to this interaction effect are shown in Fig. 7.2. These profiles indicate that the effect of muscular tension upon number of errors differs—high- and low-anxiety-level groups perform in different ways. A test on the difference between mean number of errors between the two anxiety levels under the no-tension condition (b_1) is given by

$$F = \frac{(AB_{11} - AB_{21})^2}{2nrMS_{\text{error(between)}}} = \frac{(125 - 105)^2}{2(12)(10.31)} = 1.62.$$

The critical value for this test is

$$F_{.95}(1,8) = 5.32.$$

Thus the data indicate no statistically significant difference between the high- and low-anxiety groups under the condition of no tension. A test on the difference between the high- and low-anxiety groups under the high-tension condition is given by

$$F = \frac{(AB_{12} - AB_{22})^2}{2nrMS_{\text{error(between)}}} = \frac{(104 - 146)^2}{2(12)(10.31)} = 7.13.$$

The critical value for this test is also 5.32. Hence the data indicate a statistically significant difference in the performance of the high- and low-anxiety groups under the high-tension conditions; the high-anxiety group tends to make significantly more errors than does the low-anxiety group. None of the interactions with factor *C* is statistically significant. Hence the data indicate that the shapes of the learning curves are essentially identical under each of the treatment combinations.

Variance-Covariance Matrices Associated with Numerical Example.

Apart from the homogeneity assumptions required for the validity of the usual *F* tests, the variance-covariance matrices provide information which is of use in its own right in describing processes operating in the experiment.

The variance-covariance matrices associated with the data in Table 7.4-3 are given below:

Level ab_{11}					Level ab_{12}				
	c_1	c_2	c_3	c_4		c_1	c_2	c_3	c_4
c_1	7.00	4.00	5.00	4.00	c_1	9.33	4.00	0.00	-.66
c_2		4.00	6.00	4.00	c_2		4.00	4.00	2.00
c_3			9.33	6.00	c_3			7.00	4.00
c_4				4.00	c_4				2.34

Level ab_{21}					Level ab_{22}				
	c_1	c_2	c_3	c_4		c_1	c_2	c_3	c_4
c_1	1.33	-2.00	-2.00	-1.34	c_1	3.00	3.00	1.00	-.50
c_2		4.00	2.00	1.00	c_2		4.00	2.00	.00
c_3			4.00	3.00	c_3			1.33	.33
c_4				2.34	c_4				.33

The variance-covariance matrix obtained by averaging corresponding entries in each of the above matrices is given below:

Pooled variances and covariances				
	c_1	c_2	c_3	c_4
c_1	5.16	2.25	1.00	.38
c_2		4.00	3.50	1.75
c_3			5.42	3.33
c_4				2.25

In this experiment, the levels of factor C represent successive blocks of trials in a learning experiment. Typically in this kind of experiment the variances tend to decrease as the learning increases. In the pooled variance-covariance matrix this trend is not clearly shown (5.16, 4.00, 5.42, 2.25). Further, the covariances between neighboring blocks of trials tend to be relatively higher than covariances between blocks which are farther apart. This trend is clearly shown in the pooled variance-covariance matrix. The symmetry conditions on the pooled variance-covariance matrix required for the strict validity of the usual F test do not, in general, hold for learning experiments. (Use of Hotelling's T^2 statistic would, however, provide an exact test even if the symmetry conditions do not hold.)

From the pooled covariance matrix, the average of the entries along the main diagonal defines $\overline{\text{var}}$, and the average of the entries off the main diagonal defines $\overline{\text{cov}}$. In this case, $\overline{\text{var}} = 4.21$, and $\overline{\text{cov}} = 2.04$. In terms of the latter quantities,

$$MS_{\text{Error(between)}} = \overline{\text{var}} + (r - 1) \overline{\text{cov}} = 10.33,$$

$$MS_{\text{Error(within)}} = \overline{\text{var}} - \overline{\text{cov}} = 2.17.$$

Within rounding error, these values are equal to those obtained for the corresponding mean squares in Table 7.4-6.

Illustrative Applications. The numerical example just considered is actually a modified version of an experiment reported by Meyer and Noble (1958). The plan for their experiment was a $2 \times 2 \times 6$ factorial experiment with repeated measures on the last factor. Factor *A* indicated anxiety level as measured by the Taylor Manifest Anxiety Scale. Factor *B* indicated the level of muscular tension exerted on a hand dynamometer during the experiment. Factor *C* represents blocks of trials. The criterion was the number of errors per block in the course of learning a verbal maze. The purpose of the experiment was to investigate the interaction between manifest anxiety and muscular tension during a learning task. There were 20 subjects in each of the four experimental groups.

Table 7.4-7 Analysis of Variance for Meyer and Noble Data

Source of variation	df	MS	<i>F</i>
Between subjects	79		
<i>A</i> (anxiety level)	1	2.33	
<i>B</i> (muscular tension)	1	.96	
<i>AB</i>	1	41.88	6.12*
Subj w. groups	76	6.84	
Within subjects	400		
<i>C</i> (blocks of trials)	5	216.86	309.80*
<i>AC</i>	5	.22	
<i>BC</i>	5	.29	
<i>ABC</i>	5	.58	
<i>C</i> \times subj w. groups	380	.70	

The analysis of variance reported by Meyer and Noble is given in Table 7.4-7. The significant main effect for factor *C* indicated that the mean number of errors changed during the blocks of trials. The significant *AB* interaction indicated that the effect of muscular tension is different in the two anxiety groups. Inspection of the profiles corresponding to this interaction indicated that the presence of muscular tension in the high-anxiety group tended to increase the number of errors relative to the no-tension condition. The effect in the low-anxiety group was just the opposite; under the tension condition there was a smaller number of errors relative to the no-tension condition.

A design which, on the surface, appears identical with that being considered in this section is reported by Taylor (1958). The purpose of this experiment was to study the effect of methods of prior presentation of syllables upon the threshold value for visual recognition. The plan may be represented schematically as indicated below. This method of representation, however, will be shown to be misleading.

Method of prior presentation	Association value	Lists	
		c_1	c_2
a_1	b_1	G_{11}	G_{11}
	b_2	G_{12}	G_{12}
	b_3	G_{13}	G_{13}
a_2	b_1	G_{21}	G_{21}
	b_2	G_{22}	G_{22}
	b_3	G_{23}	G_{23}

The levels of factor A define methods of prior presentation of the syllables. The levels of factor B define the association value of the syllables. However, only the syllables in list c_1 were included in the prior presentation. Syllables in list c_2 were matched with those in list c_1 for association value but were included in the experiment to serve as controls. There were 20 subjects in each group. This experimental plan does not conform to the pattern of a $2 \times 3 \times 2$ factorial experiment, since treatment combination ac_{12} is identical with treatment combination ac_{22} .

This kind of plan should not in general be analyzed as a $2 \times 3 \times 2$ factorial experiment. No meaningful interaction between factors A and C exists. One possible method of analysis is to use the difference between thresholds on syllables in lists c_1 and c_2 as the criterion measure. In this case the analysis of variance takes the following form:

Source	df
A (method)	1
B (assoc. value)	2
AB	2
Within cell	114

This type of analysis in part removes differences between subjects from experimental error. As a second phase of the analysis of variance, one may combine the two levels of factor A and treat the resulting data as a 3×2 factorial experiment having 40 subjects in each group. This analysis takes the following form.

Source	df
Between subjects	119
B (assoc. value)	2
Subj w. groups	117
Within subjects	120
C (lists)	1
BC	2
$C \times$ subj w. groups	117

Table 7.4-8 General Expected Values for Mean Squares

Effect	<i>i</i>	<i>j</i>	<i>k</i>	<i>m</i>	<i>o</i>	E(MS)
α_i	D_p	q	r	n	1	$\sigma_e^2 + D_r\sigma_{\gamma\pi}^2 + nD_qD_r\sigma_{\alpha\beta\gamma}^2 + nqD_r\sigma_{\alpha\gamma}^2 + nrD_q\sigma_{\alpha\beta}^2 + nqr\sigma_{\alpha}^2$
β_j	p	D_a	r	n	1	$\sigma_e^2 + D_r\sigma_{\gamma\pi}^2 + nD_pD_r\sigma_{\alpha\beta\gamma}^2 + npD_r\sigma_{\beta\gamma}^2 + r\sigma_{\pi}^2 + nrD_p\sigma_{\alpha\beta}^2 + npr\sigma_{\beta}^2$
$\alpha\beta_{ij}$	D_p	D_a	r	n	1	$\sigma_e^2 + D_r\sigma_{\gamma\pi}^2 + nD_r\sigma_{\alpha\beta\gamma}^2 + r\sigma_{\pi}^2 + nr\sigma_{\alpha\beta}^2$
$\pi_{m(ii)}$	1	1	r	1	1	$\sigma_e^2 + D_r\sigma_{\gamma\pi}^2 + r\sigma_{\pi}^2$
γ_k	p	q	D_r	n	1	$\sigma_e^2 + \sigma_{\gamma\pi}^2 + nD_pD_r\sigma_{\alpha\beta\gamma}^2 + npD_q\sigma_{\beta\gamma}^2 + nqD_p\sigma_{\alpha\gamma}^2 + npq\sigma_{\gamma}^2$
$\alpha\gamma_{ik}$	D_p	q	D_r	n	1	$\sigma_e^2 + \sigma_{\gamma\pi}^2 + nD_r\sigma_{\alpha\beta\gamma}^2 + nq\sigma_{\alpha\gamma}^2$
$\beta\gamma_{jk}$	p	D_a	D_r	n	1	$\sigma_e^2 + \sigma_{\gamma\pi}^2 + nD_p\sigma_{\alpha\beta\gamma}^2 + np\sigma_{\beta\gamma}^2$
$\alpha\beta\gamma_{ijk}$	D_p	D_a	D_r	n	1	$\sigma_e^2 + \sigma_{\gamma\pi}^2 + n\sigma_{\alpha\beta\gamma}^2$
$\gamma\pi_{km(ii)}$	1	1	D_r	1	1	$\sigma_e^2 + \sigma_{\gamma\pi}^2$
$\varepsilon_o(ijkm)$	1	1	1	1	1	σ_e^2

General Expected Values for Mean Squares. The expected values for the mean squares given in Table 7.4-1 are a special case of those given in Table 7.4-8. Depending upon the experimental design, D_p , D_q , and D_r are either zero or unity. The expected values in Table 7.4-1 are obtained from those in Table 7.4-8 by assuming that each of these D 's is zero, i.e., by assuming that factors A , B , and C are fixed.

7.5 Other Multifactor Repeated-measure Plans

The plans that have been considered in this chapter have the following general form:

	v_1	v_2	\cdots	v_h	
u_1	G_1	G_1	\cdots	G_1	n subjects per group
u_2	G_2	G_2	\cdots	G_2	
\vdots					
\vdots					
u_g	G_g	G_g	\cdots	G_g	

The analysis of this general form may be outlined as follows:

Source	df
Between subjects	$ng - 1$
U	$g - 1$
Subjects w. groups	$g(n - 1)$
Within subjects	$ng(h - 1)$
V	$h - 1$
UV	$(g - 1)(h - 1)$
$V \times \text{subj w. groups}$	$g(n - 1)(h - 1)$

The levels of factor U may, for example, constitute the pq treatment combinations in a $p \times q$ factorial set. In this case the following subdivisions are possible:

Source	df	Source	df
U	$g - 1$	UV	$(g - 1)(h - 1)$
A	$p - 1$	AV	$(p - 1)(h - 1)$
B	$q - 1$	BV	$(q - 1)(h - 1)$
AB	$(p - 1)(q - 1)$	ABV	$(p - 1)(q - 1)(h - 1)$

Alternatively, U may define the levels of factor A , and V may define the treatment combinations in a $q \times r$ factorial set. In this case the following subdivisions are possible:

Source	df	Source	df
V	$h - 1$	UV	$(g - 1)(h - 1)$
B	$q - 1$	AB	$(p - 1)(q - 1)$
C	$r - 1$	AC	$(p - 1)(r - 1)$
BC	$(q - 1)(r - 1)$	ABC	$(p - 1)(q - 1)(r - 1)$

Source	df
$V \times \text{subj w. groups}$	$g(n - 1)(h - 1)$
$B \times \text{subj w. groups}$	$p(n - 1)(q - 1)$
$C \times \text{subj w. groups}$	$p(n - 1)(r - 1)$
$BC \times \text{subj w. groups}$	$p(n - 1)(q - 1)(r - 1)$

Table 7.5-1 Summary of Analysis of Variance

Source of variation	df	E(MS)
Between subjects	$npq - 1$	
A	$p - 1$	$\sigma_e^2 + r\sigma_\pi^2 + nqrs\sigma_\alpha^2$
B	$q - 1$	$\sigma_e^2 + r\sigma_\pi^2 + npr\sigma_\beta^2$
AB	$(p - 1)(q - 1)$	$\sigma_e^2 + r\sigma_\pi^2 + nr\sigma_{\alpha\beta}^2$
Subjects w. groups	$pq(n - 1)$	$\sigma_e^2 + r\sigma_\pi^2$
Within subjects	$npq(rs - 1)$	
C	$r - 1$	$\sigma_e^2 + s\sigma_{\gamma\pi}^2 + npqs\sigma_\gamma^2$
AC	$(p - 1)(r - 1)$	$\sigma_e^2 + s\sigma_{\gamma\pi}^2 + nqs\sigma_{\alpha\gamma}^2$
BC	$(q - 1)(r - 1)$	$\sigma_e^2 + s\sigma_{\gamma\pi}^2 + nps\sigma_{\beta\gamma}^2$
ABC	$(p - 1)(q - 1)(r - 1)$	$\sigma_e^2 + s\sigma_{\gamma\pi}^2 + ns\sigma_{\alpha\beta\gamma}^2$
$C \times \text{subj w. groups}$	$pq(n - 1)(r - 1)$	$\sigma_e^2 + s\sigma_{\gamma\pi}^2$
D	$s - 1$	$\sigma_e^2 + r\sigma_{\delta\pi}^2 + npqr\sigma_\delta^2$
AD	$(p - 1)(s - 1)$	$\sigma_e^2 + r\sigma_{\delta\pi}^2 + nqr\sigma_{\alpha\delta}^2$
BD	$(q - 1)(s - 1)$	$\sigma_e^2 + r\sigma_{\delta\pi}^2 + npr\sigma_{\beta\delta}^2$
ABD	$(p - 1)(q - 1)(s - 1)$	$\sigma_e^2 + r\sigma_{\delta\pi}^2 + nr\sigma_{\alpha\beta\delta}^2$
$D \times \text{subj w. groups}$	$pq(n - 1)(s - 1)$	$\sigma_e^2 + r\sigma_{\delta\pi}^2$
CD	$(r - 1)(s - 1)$	$\sigma_e^2 + \sigma_{\gamma\delta\pi}^2 + npq\sigma_{\gamma\delta}^2$
ACD	$(p - 1)(r - 1)(s - 1)$	$\sigma_e^2 + \sigma_{\gamma\delta\pi}^2 + nq\sigma_{\alpha\gamma\delta}^2$
BCD	$(q - 1)(r - 1)(s - 1)$	$\sigma_e^2 + \sigma_{\gamma\delta\pi}^2 + np\sigma_{\beta\gamma\delta}^2$
$ABCD$	$(p - 1)(q - 1)(r - 1)(s - 1)$	$\sigma_e^2 + \sigma_{\gamma\delta\pi}^2 + n\sigma_{\alpha\beta\gamma\delta}^2$
$CD \times \text{subj w. groups}$	$pq(n - 1)(r - 1)(s - 1)$	$\sigma_e^2 + \sigma_{\gamma\delta\pi}^2$

As a third case, the levels of factor U may constitute pq treatment combinations in a $p \times q$ factorial set, and the levels of factor V may constitute the rs combinations in an $r \times s$ factorial set. In this case the general form specializes to the analysis summarized in Table 7.5-1. In this table, terms have been rearranged to indicate denominators for F ratios. The expected values given in this table are derived under the assumption that factors A , B , C , and D are fixed.

To illustrate the subdivision of the UV interaction for this case, one may first subdivide the U factor as follows:

Source	df
UV	$(g-1)(h-1)$
AV	$(p-1)(h-1)$
BV	$(q-1)(h-1)$
ABV	$(p-1)(q-1)(h-1)$

Then each of the interactions with factor V may be subdivided. For example, the AV interaction may be subdivided into the following parts:

Source	df
AV	$(p-1)(h-1)$
AC	$(p-1)(r-1)$
AD	$(p-1)(s-1)$
ACD	$(p-1)(r-1)(s-1)$

Analogous partitions may be made for the BV and ABV interactions.

Computational procedures for all treatment effects in the analysis summarized in Table 7.5-1 follow the usual procedures for a four-factor factorial experiment. Variation due to subjects within groups is part of the within-cell variation. The latter is given by

$$SS_{w, cell} = \sum X^2 - \frac{\sum (ABC D_{ijk\ell})^2}{n}$$

Variation due to main effects of subjects within groups is given by

$$SS_{subj w, groups} = \frac{\sum P_m^2}{rs} - \frac{\sum (AB_{ij})^2}{nrs}$$

The pooled interaction of treatment effects with subject effects is given by

$$SS_{pooled int w, subj} = SS_{w, cell} - SS_{subj w, groups}$$

This pooled interaction corresponds to the $V \times$ subjects within groups interaction in the general form of this plan. The $C \times$ subjects within groups part of this pooled interaction is given by

$$SS_{C \times subj w, groups} = \frac{\sum (CP_{km})^2}{s} - \frac{\sum (ABC_{ijk})^2}{ns} - SS_{subj w, groups}$$

Controlling Sequence Effects. For the plans that have been discussed in this chapter, in cases where the sequence of administration of the treatments was not dictated by the nature of the experimental variables, it was suggested that order be randomized independently for each subject. A partial control of sequence effects is provided by the use of the Latin-square principle; this principle is discussed in Chap. 10. A variety of repeated-measure designs using this principle is also discussed in Chap. 10. A more complete control of sequence effects (but one which is more costly in terms of experimental effort) is available. This more complete control is achieved by building what may be called a sequence factor into the design.

Consider a $p \times r$ factorial experiment in which there are to be repeated measures on the factor having r levels. The number of different sequences or arrangements of r levels is $r! = r(r-1)(r-2)(r-3) \cdots (1)$. For example, if r is 3, the number of possible sequences is $3! = 3 \cdot 2 \cdot 1 = 6$; if r is 5, the number of possible sequences is $5! = 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1 = 120$. Each of the possible sequences may define a level of factor B in which $q = r!$. Thus, instead of the original $p \times r$ factorial experiment one has a $p \times q \times r$ factorial experiment. The analysis of the latter experiment would have the following form:

Source	df
Between subjects	$npq - 1$
A	$p - 1$
B (sequence of C)	$q - 1 = r! - 1$
AB	$(p - 1)(q - 1)$
Subjects within groups	$pq(n - 1)$
Within subjects	$npq(r - 1)$
C	$r - 1$
AC	$(p - 1)(r - 1)$
BC	$(q - 1)(r - 1)$
ABC	$(p - 1)(q - 1)(r - 1)$
$C \times$ subj w. groups	$pq(n - 1)(r - 1)$

This kind of sequence factor may be constructed in connection with any repeated-measure design in which the sequence can logically be varied. However, for designs in which the number of levels of the factor on which there are repeated measures is five or more the required number of levels of the sequence factor becomes prohibitively large.

One possible method of reducing this number is to select deliberately representative sequences from among the total possible sequences. A different approach might be to take a stratified random sample of all possible sequences, where the strata are constructed so as to assure a partial balance

with respect to the order in which each of the levels appears within sequences which are used in the experiment. This kind of stratification may be achieved by using a Latin square.

7.6 Tests on Trends

Consider a $p \times q$ factorial experiment in which the levels of factor B define steps along an underlying continuum, i.e., intensity of light, dosage, blocks of trials in a learning experiment. The magnitude of the AB interaction in this kind of experiment may be regarded as a measure of global differences in the patterns or shapes of the profiles for the simple main effect of factor B . It is often of interest to study more specific aspects of such differences in patterns. Toward this end it is necessary to define dimensions in terms of which relatively irregular, experimentally determined profiles may be described. There are different methods whereby descriptive categories for this purpose may be established. In this section such categories will be defined in terms of polynomials of varying degree. Other functions, such as logarithmic or exponential, rather than polynomials, may be more appropriate for some profiles. The latter functional forms are not so readily handled as polynomials. However, polynomials may be used as first approximations to the latter forms.

Given the set of means $\overline{AB}_{i1}, \overline{AB}_{i2}, \dots, \overline{AB}_{iq}$ in a $p \times q$ factorial experiment. The line joining these means (the profile of a simple main effect of factor B at level a_i) may have an irregular shape. As a first approximation to a quantitative description of the shape of a profile, one may obtain the best-fitting linear function (straight line). The slope of this best-fitting straight line defines the linear trend of the profile. As a second approximation to the pattern of the experimentally determined set of points, one may fit a second-degree (quadratic) function. The increase in goodness of fit over the linear fit defines what is known as the quadratic trend of the profile. As a third approximation to the pattern, one may obtain the best-fitting third-degree (cubic) function. The increase in goodness of fit of the latter function over both the linear and quadratic functions defines the cubic trend of the profile.

This process of fitting polynomial functions of increasingly higher degree may be continued up to a polynomial of degree $q - 1$, where q is the number of points in the profile. A polynomial of degree $q - 1$ will always provide an exact fit to q points, since statistically there are only $q - 1$ degrees of freedom in this set of points. In most practical applications of the procedures to be described here, the degree of the polynomial is seldom carried beyond 3.

Global differences between shapes of profiles for simple main effects of factor B give rise to the AB interaction. Differences between the linear trends of such profiles define that part of the AB interaction which is called AB (linear). Thus the AB (linear) interaction represents a specific part of

the over-all AB interaction. Differences between quadratic trends in the profiles of the simple main effects define the AB (quadratic) variation. In general, the over-all variation due to AB interaction may be divided into nonoverlapping, additive parts. These parts arise from specific kinds of differences in the shapes of profiles—differences in linear trends, differences in quadratic trends, etc. Symbolically, the AB interaction may be partitioned into the following parts:

Source of variation	df
AB	$(p - 1)(q - 1)$
AB (linear)	$p - 1$
AB (quadratic)	$p - 1$
.....	
AB (degree $q - 1$)	$p - 1$

These parts will sum to the over-all AB variation.

The expected value of the mean square due to differences in linear trends has the following form:

$$E(MS_{ab(\text{lin})}) = \sigma_e^2 + n\sigma_{\alpha\beta(\text{lin})}^2.$$

A test on differences in linear trend involves the hypothesis that $\sigma_{\alpha\beta(\text{lin})}^2 = 0$. This is equivalent to a test on the hypothesis that the profiles of the simple main effects have equal slopes, i.e., that the best-fitting linear functions are parallel.

The expected value of the mean square due to differences in quadratic trends has the following general form.

$$E(MS_{ab(\text{quad})}) = \sigma_e^2 + n\sigma_{\alpha\beta(\text{quad})}^2.$$

A test on differences in these trends indicates whether or not the experimental data support the hypothesis that the profiles have equal quadratic trends.

Computational Procedures. Computational procedures will be described for the case of a $p \times q \times r$ factorial experiment in which there are repeated measures on factor C and n subjects in each group. This experimental plan has the form shown on page 355.

Suppose that the levels of factor C represent equal steps along an underlying continuum. For example, suppose that the levels of factor C are r consecutive blocks of trials in a learning experiment. Under these conditions, best-fitting linear, quadratic, cubic, etc., functions are most readily obtained by using the coefficients of orthogonal polynomials associated with r levels of an independent variable. Such coefficients are given in Table B.10.

The coefficients associated with the linear function will be designated

$$u_1, u_2, \dots, u_k, \dots, u_r.$$

		c_1	c_2	\dots	c_k	\dots	c_r
a_1	b_1	G_{11}	G_{11}		G_{11}		G
	\vdots						
	b_q	G_{1q}	G_{1q}		G_{1q}		G_{1q}
a_p	\vdots						
	b_1	G_{p1}	G_{p1}		G_{p1}		G_{p1}
	\vdots						
	b_q	G_{pq}	G_{pq}		G_{pq}		G_{pq}

For example, for the case in which $r = 4$, the respective coefficients are

$$-3, -1, 1, 3.$$

The coefficients associated with the quadratic function having r experimentally determined points will be designated

$$v_1, v_2, \dots, v_k, \dots, v_r.$$

For the case $r = 4$, the respective coefficients are

$$1, -1, -1, 1.$$

Note that the sum of the coefficients in each case is zero. Hence these coefficients define a comparison or contrast among the r points. The coefficients associated with the cubic function will be designated

$$w_1, w_2, \dots, w_k, \dots, w_r.$$

In an experiment of this kind there are three sets of interactions that may be divided into parts associated with differences between trends—the AC , the BC , and the ABC interactions. Procedures for obtaining the variation due to differences in linear trends within each of these interactions will be outlined. Higher-order trends will be found to follow the same general pattern. It will also be convenient to indicate procedures for obtaining the variation due to linear and higher-order trends for the main effect of factor C . The latter indicate the goodness of fit of polynomials of varying degree to the profile corresponding to the main effect of factor C .

The notation to be used in obtaining the linear part of the variation in interactions with factor C as well as the variation in the linear part of the variation in the main effect of factor C is summarized in part i of Table 7.6-1.

Table 7.6-1 Notation for Analysis of Linear Trend

Coefficients for linear comparison:

$$u_1, u_2, \dots, u_k, \dots, u_r$$

(i)	$X'_{m(ij)} = \sum_k u_k (X_{ijkm})$		
	$ABC'_{ij} = \sum_k u_k (ABC_{ijk}) = \sum_m X'_{m(ij)}$		
	$AC'_i = \sum_k u_k (AC_{ik}) = \sum_j ABC'_{ij}$		
	$BC'_j = \sum_k u_k (BC_{jk}) = \sum_i ABC'_{ij}$		
	$C' = \sum_k u_k C_k = \sum_i \sum_j ABC'_{ij} = \sum_i \sum_j \sum_m X'_{m(ij)}$		
(ii)	$(1') = (C')^2 / npq (\sum u_k^2)$		$(3') = \sum (AC'_i)^2 / nq (\sum u_k^2)$
	$(2') = \sum (X'_{m(ij)})^2 / (\sum u_k^2)$		$(4') = \sum (BC'_j)^2 / np (\sum u_k^2)$
	$(5') = \sum (ABC'_{ij})^2 / n (\sum u_k^2)$		
Source		Computational formula	df
Within subjects (linear)		(2')	npq
(iii)	C (linear)	(1')	1
	AC (linear)	$(3') - (1')$	$p - 1$
	BC (linear)	$(4') - (1')$	$q - 1$
	ABC (linear)	$(5') - (3') - (4') + (1')$	$(p - 1)(q - 1)$
	$C \times \text{subj w. groups}$ (linear)	$(2') - (5')$	$pq(n - 1)$

The symbol $X'_{m(ij)}$ is a weighted sum of the r observations on subject $m(ij)$, the weights being the respective linear coefficients of the appropriate polynomial. The analysis of linear trends for this case reduces to what is essentially an analysis of variance of a $p \times q$ factorial experiment with n observations per cell, an observation being an $X'_{m(ij)}$.

Other symbols defined in part i are also weighted sums. To illustrate, ABC'_{ij} is a weighted sum of terms appearing in a row of an ABC summary table. For example, the row which corresponds to level a_1 and level b_1 has the form

$$ABC_{111} \quad ABC_{112} \quad \dots \quad ABC_{11k} \quad \dots \quad ABC_{11r}$$

Each of these totals is the sum of n observations. From these totals,

$$ABC'_{ij} = u_1(ABC_{111}) + u_2(ABC_{112}) + \dots + u_r(ABC_{11r})$$

An equivalent expression for ABC'_{ij} is

$$ABC'_{ij} = X'_{1(ij)} + X'_{2(ij)} + \dots + X'_{n(ij)} = \sum_m X'_{m(ij)}$$

One expression serves as a computational check on the other.

Computational symbols convenient for use in the computation of the linear sources of variation are given in part ii. In the denominator of all

symbols is the term Σu_k^2 . The other term in the denominators is the number of observations that goes into an element which is weighted in the weighted sum. For example, there are n observations in ABC_{ijk} ; there are np observations in AC_{ik} ; there are nq observations in BC_{jk} . Actual computational formulas are summarized in part iii. The mean square due to AC (linear) estimates an expression of the form

$$\sum_i (\beta_i - \beta)^2,$$

where the β_i 's represent regression coefficients for linear profiles corresponding to simple main effects of factor C at each of the levels of factor

Table 7.6-2 Notation for Analysis of Quadratic Trend

Coefficients for quadratic comparison:	
$v_1, v_2, \dots, v_k, \dots, v_r$	
(i)	$X''_{m(ij)} = \sum_k v_k (X_{ijkm})$
	$ABC''_{ij} = \sum_k v_k (ABC_{ijk}) = \sum_m X''_{m(ij)}$
	$AC''_i = \sum_k v_k (AC_{ik}) = \sum_j ABC''_{ij}$
	$BC''_j = \sum_k v_k (BC_{jk}) = \sum_i ABC''_{ij}$
	$C'' = \sum_k v_k C_k = \sum_i \sum_j ABC''_{ij} = \sum_i \sum_j \sum_m X''_{m(ij)}$
(ii)	$(1'') = (C'')^2 / npq(\Sigma v_k^2)$
	$(2'') = \Sigma (X''_{m(ij)})^2 / (\Sigma v_k^2)$
	$(3'') = \Sigma (AC''_i)^2 / nq(\Sigma v_k^2)$
	$(4'') = \Sigma (BC''_j)^2 / np(\Sigma v_k^2)$
$(5'') = \Sigma (ABC''_{ij})^2 / n(\Sigma v_k^2)$	
Source	
Computational formula	
Within subjects (quadratic)	
(iii)	$(2'')$
	$(1'')$
	$(3'') - (1'')$
	$(4'') - (1'')$
	$(5'') - (3'') - (4'') + (1'')$
$C \times \text{subj w. groups (quadratic)}$	
$(2'') - (5'')$	

A , and where β represents a pooled regression coefficient for all linear profiles in the set.

Computational procedures for the quadratic trend are summarized in Table 7.6-2. Each of the entries in this table has a corresponding entry in Table 7.6-1, with v_k replacing corresponding u_k throughout. The symbol $X''_{m(ij)}$ is used to distinguish a weighted sum in terms of quadratic weights from the corresponding sum in terms of linear weights, designated by the symbol $X'_{m(ij)}$. Higher-order trends follow the same pattern, with

the appropriate coefficients serving as the weights. If, for example, $r = 4$, it will be found that

$$SS_e = SS_{e(\text{lin})} + SS_{e(\text{quad})} + SS_{e(\text{cubic})}.$$

Similarly, $SS_{ac} = SS_{ac(\text{lin})} + SS_{ac(\text{quad})} + SS_{ac(\text{cubic})}.$

Numerical Example. The numerical data in part i of Table 7.6-3 will be used to illustrate the computational procedures. These data represent a $2 \times 2 \times 4$ factorial experiment with repeated measures on factor C , three observations in each group. Suppose that factor C represents equally spaced blocks of trials in a learning experiment. For example, subject 1 is assigned to treatment combination ab_{11} and has scores of 1, 6, 5, and 7, respectively, on a series of four blocks of trials.

Since factor C has four levels, coefficients for the case in which there are four points to be fitted are appropriate. From Table B.10 the linear coefficients are

$$-3, -1, 1, 3.$$

These coefficients appear at the top of part i. From the data on subject 1 one obtains

$$X'_{1(11)} = (-3)(1) + (-1)(6) + (1)(5) + (3)(7) = 17,$$

the weighted sum of the observations on subject 1, the weights being the linear coefficients. Other entries in the column headed $X'_{m(ij)}$ are obtained in an analogous manner. For example,

$$X'_{4(12)} = (-3)(2) + (-1)(7) + (1)(12) + (3)(15) = 44.$$

The entries in the column headed ABC'_{ij} are obtained as follows:

$$ABC'_{11} = X'_{1(11)} + X'_{2(11)} + X'_{3(11)} = 17 + 28 + 18 = 63.$$

$$ABC'_{12} = X'_{4(12)} + X'_{5(12)} + X'_{6(12)} = 44 + 26 + 27 = 97.$$

The left-hand side of part ii represents an ABC summary table obtained in the usual manner from data on the left-hand side of part i. From the first row of the ABC summary table one obtains

$$ABC'_{11} = (-3)(4) + (-1)(20) + (1)(20) + (3)(25) = 63.$$

This entry provides a check on the entry ABC'_{11} computed in part i. The second row of the ABC summary table is used to obtain ABC'_{12} . A corresponding entry is available from part i.

The left-hand side of part iii represents an AC summary table. The entries in the column headed AC'_i are weighted sums of the entries in the corresponding rows. Checks on these entries may be obtained from part ii. Checks on entries in part iv may also be obtained from the ABC'_{ij} column of part ii. For example,

$$BC'_1 = ABC'_{11} + ABC'_{21} = 63 + 92 = 155;$$

$$BC'_2 = ABC'_{12} + ABC'_{22} = 97 + 118 = 215.$$

Table 7.6-3 Analysis of Linear Trend—Numerical Example

Linear coefficients: -3 -1 1 3 $\Sigma u_k^2 = 20$									
		Subject	c_1	c_2	c_3	c_4	$X'_{m(ij)}$	ABC'_{ij}	
(i)	a_1	1 b_1 2 3	1	6	5	7	17	63	
			0	6	7	9	28		
			3	8	8	9	18		
		4 b_2 5 6	2	7	12	15	44		97
			1	6	8	9	26		
			3	7	10	11	27		
	a_2	7 b_1 8 9	1	2	7	12	38	92	
			1	1	4	10	30		
			1	1	4	8	24		
		10 b_2 11 12	2	2	8	12	36		118
			3	2	10	15	44		
			2	2	7	13	38		
		20	50	90	130	370	370		

		c_1	c_2	c_3	c_4	ABC'_{ij}	AC'_i		
(ii)	a_1	b_1	4	20	20	25	63	160	
		b_2	6	20	30	35	97		
	a_2	b_1	3	4	15	30	92	210	
			7	6	25	40	118		
		b_2	20	50	90	130	370		370

		c_1	c_2	c_3	c_4	AC'_i	C'
(iii)	a_1 a_2	10	40	50	60	160	370
		10	10	40	70	210	
		20	50	90	130	370	

		c_1	c_2	c_3	c_4	BC'_j	C'
(iv)	b_1 b_2	7	24	35	55	155	370
		13	26	55	75	215	
		20	50	90	130	370	

All the totals required in the analysis of linear trend may be obtained from the $X'_{m(ij)}$ column of part i of Table 7.6-3. Hence parts ii to iv are not actually required. In practice, however, part ii should be computed to serve as a check on the $X'_{m(ij)}$ column. Additional checks are provided by parts iii and iv.

Computational symbols defined in part ii of Table 7.6-1 are obtained in part i of Table 7.6-4. For the case in which there are four points in each profile, $\sum u_k^2 = 20$. Each total C_k is the sum of $npq = 12$ observations. Hence the denominator for symbol (1') is 12(20). Since each total AC_{ik} is the sum of $nq = 6$ observations, the denominator for the symbol (3') is 6(20).

Table 7.6-4 Analysis of Linear Trend—Numerical Example

		(1') = $(370)^2/12(20)$	= 570.42		
		(2') = $(17^2 + 28^2 + \dots + 44^2 + 38^2)/20$	= 616.70		
(i)		(3') = $(160^2 + 210^2)/6(20)$	= 580.83		
		(4') = $(155^2 + 215^2)/6(20)$	= 585.42		
		(5') = $(63^2 + 97^2 + 92^2 + 118^2)/3(20)$	= 596.10		
Source of variation		SS	df	MS	F
Within subjects (linear)		(2') = 616.70	12		
(ii)	C (linear)	(1') = 570.42	1	570.42	221.09*
	AC (linear)	(3') - (1') = 10.41	1	10.41	4.03
	BC (linear)	(4') - (1') = 15.00	1	15.00	5.81*
	ABC (linear)	(5') - (3') - (4') + (1') = 0.27	1	0.27	
	C \times subj w. groups (linear)	(2') - (5') = 20.60	8	2.58	

The analysis of variance for the linear trend of the within-subject effects is summarized in part ii of Table 7.6-4. For the basic data in part i of Table 7.6-3, there are $npq = 12$ subjects and $r = 4$ observations on each subject. Hence the total degrees of freedom for within-subject effects are $npq(r - 1) = 36$. In the analysis of variance of trend, these 36 degrees of freedom are partitioned as follows:

Within subjects	36	$npq(r - 1)$
Within subjects (linear)	12	npq
Within subjects (quadratic)	12	npq
Within subjects (cubic)	12	npq

The 12 degrees of freedom for the linear trend of the within-subject effects are analyzed in part ii of Table 7.6-4.

A test on linear trend in the main effect of factor C has the form

$$F = \frac{MS_{e(\text{lin})}}{MS_{C \times \text{subj w. groups}(\text{lin})}} = 221.09.$$

At the .05 level of significance, this test indicates that the best-fitting straight line to the profile of the C main effects has a slope which is significantly different from zero. The profile corresponding to the C main effects (in terms of treatment means) is shown in Fig. 7.3.

The profiles corresponding to the simple effects of factor C at levels b_1 and b_2 are also shown in this figure. Inspection of these profiles suggests that the best-fitting line to the profile for b_2 would have a different slope from the best-fitting line to the profile for b_1 , that is, that these lines would not be

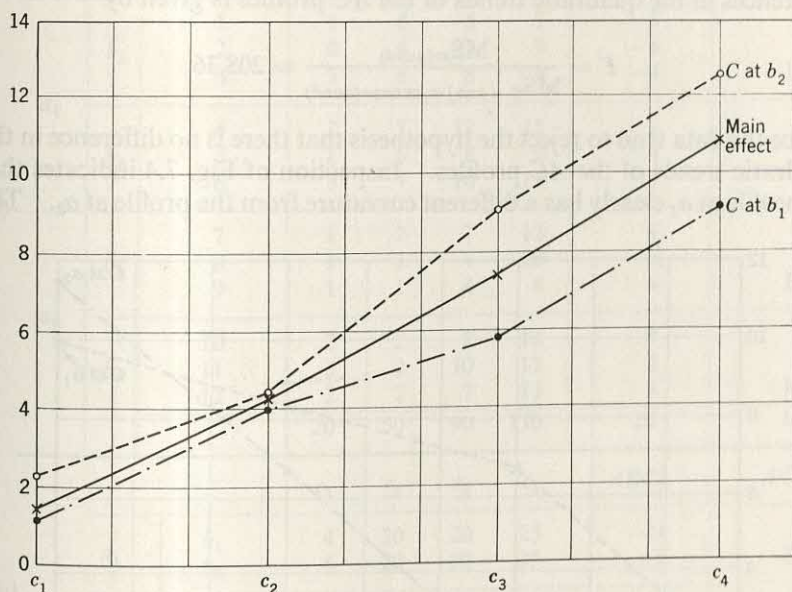


Figure 7.3 Profiles of BC interaction and C main effect.

parallel. A test of the hypothesis that there is no difference in these slopes (no difference in linear trend) has the form

$$F = \frac{MS_{bc(\text{lin})}}{MS_{C \times \text{subj w. groups}(\text{lin})}} = 5.81.$$

At the .05 level of significance, this test indicates that the linear trends of the BC profiles cannot be considered to be equal.

Profiles corresponding to the AC interaction are shown in Fig. 7.4. Inspection indicates that these profiles differ in shape but that the best-fitting straight lines might be parallel. The test of the latter hypothesis is given by

$$F = \frac{MS_{ac(\text{lin})}}{MS_{C \times \text{subj w. groups}(\text{lin})}} = 4.03.$$

The value does not exceed the critical value for a .05-level test. Hence the experimental evidence does not reject the hypothesis that the linear trends are equal. Thus differences in shapes of these profiles, if there are statistically significant differences, must be due to quadratic or higher-order trends.

A summary of the analysis of variance for the quadratic trend is given in Table 7.6-5 and Table 7.6-6. This analysis follows the same general procedures as those used in the analysis of the linear trend; in this case the quadratic coefficients replace the linear coefficients as weights. The test of differences in the quadratic trends of the AC profiles is given by

$$F = \frac{MS_{ac(quad)}}{MS_{C \times \text{subj w. groups}(quad)}} = 208.36.$$

Hence the data tend to reject the hypothesis that there is no difference in the quadratic trends of the AC profiles. Inspection of Fig. 7.4 indicates that the profile at a_1 clearly has a different curvature from the profile at a_2 . The

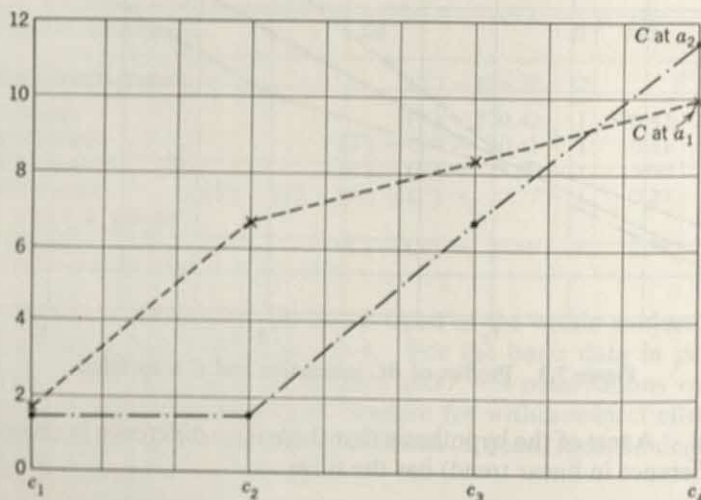


Figure 7.4 Profiles of AC interaction.

significant F ratio for the main effect in Table 7.6-6 indicates that the quadratic trend in the profile of the C main effect is different from zero.

In general, the analysis of differences in trends for interaction terms is not made unless there is evidence to show that some difference in shapes exists. This evidence is provided by the usual over-all tests on interactions, which indicate the presence or absence of global differences in the shapes of profiles. The over-all analysis of variance for the within-subject effects is summarized in Table 7.6-8. This analysis is made by means of the usual computational procedures for a $p \times q \times r$ factorial experiment having repeated measures on factor C . The basic observational data in part i of Table 7.6-3 and the summary tables at the left of parts ii to iv are used in the over-all analysis of variance. The significant AC and BC interactions indicate that there are global differences in the trends of the corresponding

Table 7.6-5 Analysis of Quadratic Trend—Numerical Example

Quadratic coefficients			1	-1	-1	1	$\Sigma v_k^2 = 4$		
			Subject	c_1	c_2	c_3	c_4	$X_{m(i)}^*$	ABC_{ij}^*
(i)	a_1	b_1	1	1	6	5	7	-3	-11
			2	0	6	7	9	-4	
			3	3	8	8	9	-4	
	b_2	4	2	7	12	15	-2	-9	
		5	1	6	8	9	-4		
		6	3	7	10	11	-3		
a_2	b_1	7	1	2	7	12	4	14	
		8	1	1	4	10	6		
		9	1	1	4	8	4		
	b_2	10	2	2	8	12	4	16	
		11	3	2	10	15	6		
		12	2	2	7	13	6		
			20	50	90	130	10	10	

(ii)			c_1	c_2	c_3	c_4	ABC_{ij}^*	AC_i^*
	a_1	b_1	4	20	20	25	-11	-20
		b_2	6	20	30	35	-9	
	a_2	b_1	3	4	15	30	14	30
			7	6	25	40	16	
		b_2	20	50	90	130	10	10

Table 7.6-6 Analysis of Quadratic Trend—Numerical Example

(i)	$(1^*) = (10)^2/12(4)$	$= 2.08$			
	$(2^*) = (-3)^2 + (-4)^2 + \dots + 6^2 + 6^2$	$= 56.50$			
	$(3^*) = [(-20)^2 + 30^2]/6(4)$	$= 54.17$			
	$(4^*) = [3^2 + 7^2]/6(4)$	$= 2.42$			
	$(5^*) = [(-11)^2 + (-9)^2 + 14^2 + 16^2]/3(4)$	$= 54.50$			
Source of variation		SS	df	MS	F
Within subjects (quadratic)		$(2^*) = 56.51$	12		
C (quadratic)		$(1^*) = 2.08$	1	2.08	8.32* 208.36*
AC (quadratic)		$(3^*) - (1^*) = 52.09$	1	52.09	
BC (quadratic)		$(4^*) - (1^*) = 0.34$	1	0.34	
(ii) ABC (quadratic)		$(5^*) - (3^*)$			
		$-(4^*) + (1^*) = 0.00$	1	0.00	
C \times subj. w. groups (quadratic)		$(2^*) - (5^*) = 2.00$	8	0.25	

Table 7.6-7 Analysis of Cubic Trend—Numerical Example

Source of variation	SS	df	MS	F
Within subject (cubic)	20.30	12		
<i>C</i> (cubic)	0.42	1	0.42	
<i>AC</i> (cubic)	10.42	1	10.42	30.64*
<i>BC</i> (cubic)	6.66	1	6.66	19.59*
<i>ABC</i> (cubic)	0.07	1	0.07	
<i>C</i> × subj w. groups (cubic)	2.73	8	0.34	

Table 7.6-8 Over-all Analysis of Variance for Within Subjects Effects—Numerical Example

Source of variation	SS	df	MS	F
Within subjects	693.50	36		
<i>C</i>	572.92	3	190.97	180.16*
<i>AC</i>	72.92	3	24.30	22.92*
<i>BC</i>	22.00	3	7.33	6.92*
<i>ABC</i>	0.33	3	0.11	
<i>C</i> × subj w. groups	25.34	24	1.06	

profiles. The nature of such differences is explored in the analysis of the linear, quadratic, and cubic trends.

It is of interest to compare the over-all analysis with the analysis of the individual trends. It will be noted that:

	SS	df
Within subjects (over-all)	693.50	36
Within subjects (linear)	616.70	12
Within subjects (quadratic)	56.51	12
Within subjects (cubic)	20.30	12

It will also be noted that:

	SS
<i>AC</i> (over-all)	72.92
<i>AC</i> (linear)	10.41
<i>AC</i> (quadratic)	52.09
<i>AC</i> (cubic)	10.42

In each case the sum of the parts will be numerically equal to the corresponding over-all variation.

It is of particular interest to look at the parts of terms that go into the denominator of the F ratios. In this case:

	SS	MS
$C \times \text{subj w. groups (over-all)}$	25.34	1.06
$C \times \text{subj w. groups (linear)}$	20.60	2.58
$C \times \text{subj w. groups (quadratic)}$	2.00	0.25
$C \times \text{subj w. groups (cubic)}$	2.73	0.34

The parts of the over-all variation of this interaction need not be homogeneous; each of these parts does not estimate the same source of variation. The parts measure the deviation of a weighted sum about the mean of the weighted sums; the weights are different for each of the parts. The expected values of the mean squares of the parts are to some extent dependent upon the weights; the latter in turn determine the shape of the curve that is being fitted to the experimental data. Different structural models underlie the analysis of variance for linear, quadratic, and cubic trends.

Illustrative Applications. Grant (1956) gives a relatively complete account of tests for trends as well as a detailed numerical example. The plan for the experiment discussed by Grant has the following form:

		Shock	Stages of task				
			c_1	c_2	c_3	c_4	c_5
Anxiety level	a_1	b_1	G_{11}	G_{11}	G_{11}	G_{11}	G_{11}
		b_2	G_{12}	G_{12}	G_{12}	G_{12}	G_{12}
		b_3	G_{13}	G_{13}	G_{13}	G_{13}	G_{13}
	a_2	b_1	G_{21}	G_{21}	G_{21}	G_{21}	G_{21}
		b_2	G_{22}	G_{22}	G_{22}	G_{22}	G_{22}
		b_3	G_{23}	G_{23}	G_{23}	G_{23}	G_{23}

Factor A represents the level of anxiety of the subjects as measured by the Taylor Manifest Anxiety Scale. Groups G_{11} , G_{12} , and G_{13} represent subjects from one end of the scale; groups G_{21} , G_{22} , and G_{23} represent subjects from the other end of the scale. The three levels of factor B indicate the number of electric shocks received by a subject in the course of the experiment. Subjects in groups G_{11} and G_{21} were at level b_1 , subjects in groups G_{12} and G_{22} at level b_2 , and subjects in groups G_{13} and G_{23} at level b_3 . The levels of factor C represent successive stages of a card-sorting task. There were four subjects in each group. The criterion was the square root of the number of perseverative errors. This plan may be considered as a special case of a $p \times q \times r$ factorial experiment with repeated measures on factor C .

The over-all analysis, as given by Grant, had the following form:

	SS	df	MS	<i>F</i>
Between subjects		23		
Groups	21.91	5	4.38	2.10
Subj w. groups	37.56	18	2.09	
Within subjects		96		
<i>C</i> stages	16.37	4	4.09	5.37*
Stages \times groups	35.26	20	1.76	2.31
Stages \times subj w. groups	54.89	72	0.76	

In this analysis the six combinations of anxiety and shock are considered to be six levels of a single factor—a “group” factor. The group factor is then subdivided as follows:

	SS	df	MS	<i>F</i>
Groups	21.91	5		
<i>A</i> anxiety level	1.75	1	1.75	0.84
<i>B</i> shock	8.86	2	4.43	2.12
<i>AB</i>	11.29	2	5.65	2.71

The denominator for each of the above *F* ratios is $MS_{\text{subj w. groups}}$.

The 20 degrees of freedom for the stage \times group interaction are associated with what Grant calls the between-group trends. This interaction is subdivided into the following parts:

	SS	df	MS	<i>F</i>
Stages \times groups	35.26	20		
<i>AC</i>	17.58	4	4.40	5.79
<i>BC</i>	10.62	8	1.33	1.75
<i>ABC</i>	7.07	8	0.88	1.16

The denominator for each of the above *F* ratios is $MS_{\text{stages} \times \text{subj w. groups}}$. These tests indicate whether or not there are any global differences in the shapes of corresponding profiles.

The more specific analysis of differences in trends associated with factor *C* is summarized in Table 7.6-9. In each case the denominator for tests on differences in linear trend is the linear part of the *C* \times subjects within groups interaction. Similarly, the quadratic part of this latter interaction serves as the denominator for differences in quadratic trend.

It will be noted that, although the over-all test on the BC interaction indicates no significant difference, the test on differences in linear trends does indicate a statistically significant difference. In the absence of a significant over-all interaction the experimenter would not ordinarily make tests on the parts—unless a priori information about the underlying sources of variation in the experimental variables indicates that certain of the trends

Table 7.6-9 Analysis of Variance of Grant Data (1956)

Source of variation	SS	df	MS	<i>F</i>
Within subjects		96		
...		...		
<i>AC</i>	17.57	4	4.39	5.78*
Linear	16.38	1	16.38	37.99*
Quadratic	0.04	1	0.04	0.05
Cubic	0.29	1	0.29	0.26
Quartic	0.86	1	0.86	1.52
<i>BC</i>	10.62	8	1.33	1.75
Linear	6.84	2	3.42	7.93*
Quadratic	1.00	2	0.50	0.55
Cubic	2.29	2	1.14	1.00
Quartic	0.49	2	0.25	0.44
<i>ABC</i>	7.07	8	0.88	1.16
Linear	2.05	2	1.02	2.37
Quadratic	1.48	2	0.74	0.81
Cubic	0.24	2	0.12	0.10
Quartic	3.30	2	1.65	2.93
<i>C</i> × subj w. groups	54.89	72	0.76	
Linear	7.76	18	0.43	
Quadratic	16.43	18	0.91	
Cubic	20.55	18	1.14	
Quartic	10.15	18	0.56	

should be more dominant than others. However, one should not hesitate to present a complete description of the experimental findings, even though some of the “tests” on parts of nonsignificant over-all variation may be unduly subject to type I error.

Another illustrative example of the analysis of trends is given in an experiment reported by Schrier (1958). The plan for this experiment may be represented as shown on page 369. In this design there are repeated measures on both factors A and B . The two levels of factor A represent two blocks of trials. The four levels of factor B represent four amounts of reward. The criterion measure was the proportion of correct discriminations. To obtain homogeneity of variance, an arcsine transformation on the proportions was used in the analysis.

Table 7.6-10 Analysis of Variance of Linear Trend

Linear coefficients: $u_1, u_2, \dots, u_j, \dots, u_q$

(i)	$X'_{im} = \sum_j u_j X_{ijm}$	$AB'_i = \sum_m X'_{im}$
	$P'_m = \sum_i X'_{im}$	$B' = \sum_m P'_m = \sum_i AB'_i$
(ii)	$(1') = (B')^2 / np(\sum u_j^2)$	$(3') = \Sigma (AB'_i)^2 / n(\sum u_j^2)$
	$(2') = \Sigma (X'_{im})^2 / (\sum u_j^2)$	$(4') = \Sigma (P'_m)^2 / p(\sum u_j^2)$
Source of variation		df
(iii)	B (linear)	(1')
	$B \times$ subjects (linear)	(4') - (1')
	AB (linear)	(3') - (1')
	$AB \times$ subjects (linear)	(2') - (3') - (4') + (1')
		1
		$n - 1$
		$p - 1$
		$(n - 1)(p - 1)$

Table 7.6-11 Analysis of Variance for Schrier Data (1958)

Source of variation		df	MS	F	Denominator
Between subjects		4	117.8		
Within subjects		35			
U	Blocks	1	379.8	23.2*	U
	Blocks \times subjects	4	16.3		
V	Rewards	3	446.1	16.2*	V
	Linear	1	1290.5	18.5*	V ₁
	Quadratic	1	1.5		V ₂
	Cubic	1	46.1	5.7*	V ₃
	Rewards \times subjects	12	27.6		
V ₁	Linear	4	69.6		
V ₂	Quadratic	4	4.6		
V ₃	Cubic	4	8.2		
W	Blocks \times rewards	3	41.2	6.2*	W
	Linear	1	75.0	6.6	W ₁
	Quadratic	1	.2		W ₂
	Cubic	1	48.6	16.6*	W ₃
W	Blocks \times rewards \times subjects	12	6.6		
	Linear	4	11.5		
	Quadratic	4	5.6		
	Cubic	4	3.0		

Subjects	Trial: a_1					a_2			
	Reward level:	b_1	b_2	b_3	b_4	b_1	b_2	b_3	b_4
1									
2									
.									
.									
5									

Since there are repeated measures on two factors, computational procedures for this design differ somewhat from those given in Table 7.6-1. General procedures for the analysis of linear trends with respect to factor B are outlined in Table 7.6-10. If the logic of the design permits, a trend analysis may also be carried out with respect to factor A . (In the case of the Schrier plan factor A has only two levels.) Higher-order trends may be analyzed by following the general pattern in Table 7.6-10, replacing the linear coefficients with appropriate higher-order coefficients throughout.

The analysis of variance reported by Schrier is summarized in Table 7.6-11. Denominators for F ratios are indicated in the right-hand column. For example, the denominator used in the test for differences in cubic trend of the blocks \times rewards interaction is the entry in row W_3 . The latter is $MS_{\text{blocks} \times \text{rewards} \times \text{subj(cubic)}}$. The experimental data indicate that the differences in cubic trend in the block \times reward interaction cannot be considered to be zero; i.e., the shapes of corresponding profiles do not have equal cubic curvature.

7.7 Testing Equality and Symmetry of Covariance Matrices

Consider the following $p \times q$ factorial experiment in which there are repeated measures on factor B .

	b_1	b_2	\cdots	b_q
a_1	G_1	G_1	\cdots	G_1
a_2	G_2	G_2	\cdots	G_2
.	.	.		.
.	.	.		.
a_p	G_p	G_p	\cdots	G_p

Assume that there are n_1 subjects in the group (G_1) assigned to level a_1 , n_2 subjects in the group (G_2) of subjects assigned to level a_2 , etc. Let $N = n_1 + n_2 + \cdots + n_p$.

The following additional notation will be used in this section:

$S_1 = q \times q$ matrix of covariances for level a_1 .

$S_2 = q \times q$ matrix of covariances for level a_2 .

.....

$S_p = q \times q$ matrix of covariances for level a_p .

$S_{\text{pooled}} = q \times q$ matrix of pooled covariances; i.e., each entry is a weighted average of corresponding entries in S_1 through S_p , the weights being the corresponding degrees of freedom.

The variables included in the covariance matrix at level a_i are assumed to have an underlying q -variate normal distribution.

The test procedures to be described in this section are those suggested by Box (1950). To test the hypothesis that the covariance matrices S_1, S_2, \dots, S_p are random samples for populations in which the covariance matrices are $\Sigma_1 = \Sigma_2 = \dots = \Sigma_p = \Sigma$ (that is, that the population covariance matrices are equal), one computes the following statistics:

$$(1) \quad M_1 = N \ln |S_{\text{pooled}}| - \sum n_i \ln |S_i|,$$

$$(2) \quad C_1 = \frac{2q^2 + 3q - 1}{6(q+1)(p-1)} \left[\sum \left(\frac{1}{n_i} \right) - \frac{1}{N} \right],$$

$$(3) \quad f_1 = \frac{q(q+1)(p-1)}{2}.$$

Under the hypothesis that the multivariate normal populations have equal covariance matrices, the statistic

$$(4) \quad \chi_1^2 = (1 - C_1)M_1$$

has a sampling distribution which is approximated by a chi-square distribution having f_1 degrees of freedom. Rejection of this hypothesis rules against pooling covariance matrices. If the populations have a common covariance matrix Σ , then S_{pooled} is an unbiased estimate of Σ . This test procedure is a multivariate analogue of Bartlett's test for homogeneity of variance. Its power is adequate only if each n_i is large relative to q .

The model under which the usual F tests in a repeated-measure factorial experiment are valid not only assumes that the matrix of covariances within

each of the populations is Σ (the same for each of the populations) but also that Σ has the following form:

$$\Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}.$$

That is, each entry on the main diagonal is equal to σ^2 , and each entry off the main diagonal is equal to $\rho\sigma^2$. If, in fact, Σ has this form, then the matrix

$$S_0 = \begin{bmatrix} \overline{\text{var}} & \overline{\text{cov}} & \cdots & \overline{\text{cov}} \\ \overline{\text{cov}} & \overline{\text{var}} & \cdots & \overline{\text{cov}} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \overline{\text{cov}} & \overline{\text{cov}} & \cdots & \overline{\text{var}} \end{bmatrix},$$

where $\overline{\text{var}}$ = mean of entries on main diagonal of S_{pooled} ,

$\overline{\text{cov}}$ = mean of entries off main diagonal of S_{pooled} ,

provides an unbiased estimate of Σ .

To test the hypothesis that Σ has the form given above, one computes the following statistics:

$$(5) \quad M_2 = -(N - p) \ln \frac{|S_{\text{pooled}}|}{|S_0|},$$

$$(6) \quad C_2 = \frac{q(q+1)^2(2q-3)}{6(N-p)(q-1)(q^2+q-4)},$$

$$(7) \quad f_2 = \frac{q^2 + q - 4}{2}.$$

Under the hypothesis that Σ has the specified form, the statistic

$$(8) \quad \chi_2^2 = (1 - C_2)M_2$$

has a sampling distribution which can be approximated by a chi-square distribution having f_2 degrees of freedom.

The computation of the statistics defined in (4) and (8) will be illustrated through use of the numerical data given in Table 7.7-1. In this table, $p = 2$, $q = 3$, and $n_1 = n_2 = 5$. The covariance matrices obtained from the data in part i of this table are given at the left in Table 7.7-2. For example, the entry 1.75 in the covariance matrix at level a_1 is the covariance between the observations in columns b_1 and b_2 at level a_1 . (The symbols in part iii of Table 7.7-1 are defined in Table 7.2-3.)

Table 7.7-1 Numerical Example

	Subject	b_1	b_2	b_3	Total	
(i)	a_1	1	4	7	2	13
		2	3	5	1	9
		3	7	9	6	22
		4	6	6	2	14
		5	5	5	1	11
		25	32	12	69	
	a_2	6	8	2	5	15
7		4	1	1	6	
8		6	3	4	13	
9		9	5	2	16	
10		7	1	1	9	
34		12	13	59	128 = G	

(ii)		b_1	b_2	b_3	
	a_1	25	32	12	69
	a_2	34	12	13	59
		59	44	25	128

(iii)	$(2a_1) = 397$	$(2a_2) = 333$
	$(3a_1) = 317.40$	$(3a_2) = 232.07$
	$(5a_1) = 358.60$	$(5a_2) = 293.80$
	$(6a_1) = 350.33$	$(6a_2) = 255.67$

(iv)	$SS_{B \times \text{subj w. } a_1} = (2a_1) - (5a_1) - (6a_1) + (3a_1) = 5.47$
	$SS_{B \times \text{subj w. } a_2} = (2a_2) - (5a_2) - (6a_2) + (3a_2) = 15.60$
	$SS_{B \times \text{subj (pooled)}} = 21.07$
	$MS_{B \times \text{subj w. } a_1} = 5.47/8 = 0.684$
	$MS_{B \times \text{subj w. } a_2} = 15.60/8 = 1.950$
	$MS_{B \times \text{subj (pooled)}} = 21.04/16 = 1.315$

Table 7.7-2 Covariance Matrices Associated with Data in Table 7.7-1

	a_1				a_2				<i>Pooled</i>		
	b_1	b_2	b_3		b_1	b_2	b_3		b_1	b_2	b_3
b_1	2.50	1.75	2.50	b_1	3.70	2.10	1.15	b_1	3.10	1.92	1.82
b_2	1.75	2.80	3.30	b_2	2.10	2.80	0.70	b_2	1.92	2.80	2.00
b_3	2.50	3.30	4.30	b_3	1.15	0.70	3.30	b_3	1.82	2.00	3.80
$\overline{\text{var}} - \overline{\text{cov}} = 0.68$			$\overline{\text{var}} - \overline{\text{cov}} = 1.95$			$\overline{\text{var}} - \overline{\text{cov}} = 1.32$					

The matrix corresponding to S_{pooled} is given at the right of Table 7.7-2. The entry 3.10 is the mean of corresponding entries at levels a_1 and a_2 , that is, $(2.50 + 3.70)/2 = 3.10$. Similarly, the entry 1.92 is the mean of corresponding entries at a_1 and a_2 . The numerical values of the determinants corresponding to S_1 , S_2 , and S_{pooled} are

$$|S_1| = 1.08, \quad |S_2| = 17.51, \quad |S_{\text{pooled}}| = 11.28.$$

The statistic defined in (4) is obtained from

$$\begin{aligned} M_1 &= 10 \ln(11.28) - 5 \ln(1.08) - 5 \ln(17.51) \\ &= 10(2.421) - 5(0.077) - 5(2.863) \\ &= 9.510, \end{aligned}$$

$$C_1 = \frac{18 + 9 - 1}{6(4)(1)} \left(\frac{1}{5} + \frac{1}{5} - \frac{1}{10} \right) = .325,$$

$$f_1 = \frac{3(4)(1)}{2} = 6.$$

Hence

$$\chi_1^2 = (1 - .325)(9.510) = 6.42.$$

If $\alpha = .05$, the critical value for the test of homogeneity of population covariances matrices is $\chi_{.95}^2(6) = 12.6$. Since the observed chi-square statistic does not exceed the latter value, the hypothesis of homogeneity of covariances may be considered tenable. An unbiased estimate of Σ is given by S_{pooled} .

The matrix S_0 is given by

$$S_0 = \begin{bmatrix} 3.23 & 1.91 & 1.91 \\ 1.91 & 3.23 & 1.91 \\ 1.91 & 1.91 & 3.23 \end{bmatrix}.$$

Note that

$$\overline{\text{var}} - \overline{\text{cov}} = 3.23 - 1.91 = 1.32.$$

Within rounding error, this is the value of $MS_{B \times \text{subj}}$ computed in part iv of Table 7.7-1. The numerical value of the determinant corresponding to S_0 is

$$|S_0| = 12.30.$$

To obtain the statistic defined in (8),

$$M_2 = -(10 - 3) \ln \frac{11.28}{12.30} = -7(-.0866) = .692,$$

$$C_2 = \frac{3(4)(3)}{6(8)(2)(8)} = 0.47.$$

$$f_2 = \frac{9 + 1 - 4}{2} = 3.$$

Hence

$$\chi_2^2 = (1 - .047)(.692) = .659.$$

The critical value for a .05-level test on the hypothesis that all the diagonal values of Σ are σ^2 and all the off-diagonal entries are $\rho\sigma^2$ is $\chi^2_{.95}(3) = 7.8$. These data do not contradict this hypothesis.

It is of interest to note the following relationships:

$$MS_{\text{subj w. } a_1} = 8.23,$$

$$\overline{\text{var}}_{a_1} + (q - 1) \overline{\text{cov}}_{a_1} = 3.20 + (2)(2.52) = 8.24;$$

$$MS_{\text{subj w. } a_2} = 5.90,$$

$$\overline{\text{var}}_{a_2} + (q - 1) \overline{\text{cov}}_{a_2} = 3.27 + (2)(1.32) = 5.91.$$

In general, within rounding error,

$$MS_{\text{subj w. } a_i} = \overline{\text{var}}_{a_i} + (q - 1) \overline{\text{cov}}_{a_i}.$$

7.8 Unequal Group Size

Consider the following design:

	c_1		\dots	c_r		Group size
	b_1	\dots b_q		b_1	\dots b_q	
a_1	G_1	G_1		G_1	G_1	n_1
a_2	G_2	G_2		G_2	G_2	n_2
\vdots						
a_p	G_p	G_p		G_p	G_p	n_p

Total number of subjects = N

If the levels of factor A represent different strata within a specified population, then the n_i may be proportional to the number of individuals actually in each of these strata in the population. In this case, a least-squares solution for the effects and the sums of squares is appropriate. However, if the original plan for an experiment calls for equal group size, but the completed experiment does not have equal group size because of conditions unrelated to the treatments per se, then an unweighted-means solution is the more appropriate. Both types of solution are considered in this section.

Because the cell frequencies will be proportional by columns, the least-squares solution can be obtained quite simply. Computational symbols for this case are defined in Table 7.8-1. The sums of squares are obtained by the relations given in Table 7.3-6, using the symbols as defined in Table 7.8-1. The degrees of freedom for the sums of squares are obtained from

Table 7.8-1 Unequal Group Size—Least-squares Solution

(1) $= \frac{G^2}{Nqr}$	(6) $= \frac{\Sigma[(AB_{ij})^2/n_i]}{r}$
(2) $= \Sigma X^2$	(7) $= \frac{\Sigma[(AC_{ik})^2/n_i]}{q}$
(3) $= \frac{\Sigma(A_i^2/n_i)}{qr}$	(8) $= \frac{\Sigma(BC_{jk})^2}{N}$
(4) $= \frac{\Sigma B_j^2}{Nr}$	(9) $= \Sigma \left[\frac{(ABC_{ijk})^2}{n_i} \right]$
(5) $= \frac{\Sigma C_k^2}{Nq}$	(10) $= \frac{\Sigma P_m^2}{qr}$

those given in Table 7.3-1 by replacing np with N throughout, where $N = \Sigma n_i$. For example,

$$\begin{array}{ll} p(n-1) & \text{becomes} \quad N-p, \\ p(n-1)(q-1) & \text{becomes} \quad (N-p)(q-1). \end{array}$$

The starting point for an unweighted-means solution is an ABC' summary table in which a cell entry is a mean; that is,

$$ABC'_{ijk} = \frac{ABC_{ijk}}{n_i}.$$

From this summary table, AB' and AC' summaries are computed in the usual manner; that is,

$$\begin{aligned} AB'_{ij} &= \sum_k ABC'_{ijk}, \\ A'_i &= \sum_j AB'_{ij}. \end{aligned}$$

Computational symbols appropriate for this case are given in Table 7.8-2. Only those sums of squares which do not involve the subject factor are

Table 7.8-2 Unequal Group Size—Unweighted-means Solution

(i)	(1') $= G^2/pqr$	(6') $= [\Sigma(AB'_{ij})^2]/r$
	(3') $= (\Sigma A_i'^2)/qr$	(7') $= [\Sigma(AC'_{ik})^2]/q$
	(4') $= (\Sigma B_j^2)/pr$	(8') $= [\Sigma(BC'_{jk})^2]/p$
	(5') $= (\Sigma C_k^2)/pq$	(9') $= \Sigma(ABC'_{ijk})^2$
(ii)	$SS_a = \tilde{n}_h[(3') - (1')]$	$SS_{ab} = \tilde{n}_h[(6') - (3') - (4') + (1')]$
	$SS_b = \tilde{n}_h[(4') - (1')]$	$SS_{ac} = \tilde{n}_h[(7') - (3') - (5') + (1')]$
	$SS_c = \tilde{n}_h[(5') - (1')]$	$SS_{bc} = \tilde{n}_h[(8') - (4') - (5') + (1')]$
	$SS_{abc} = \tilde{n}_h[(9') - (6') - (7') - (8') + (3') + (4') + (5') - (1')]$	

**Table 7.8-5 Analysis of Variance for Numerical Example—
Unweighted-means Solution**

Source of variation	Computational formula	SS	df	MS	F
Between subjects			<u>11</u>		
<i>A</i>	$\bar{n}_h[(3') - (1')]$	23.17	2	11.29	
Subjects w. groups	$(6) - (3)$	612.62	9	68.06	
Within subjects			<u>24</u>		
<i>B</i>	$\bar{n}_h[(4') - (1')]$	346.73	2	173.36	79.89
<i>AB</i>	$\bar{n}_h[(5') - (3') - (4') + (1')]$	179.86	4	44.97	20.72
<i>B</i> × subjects w. groups	$(2) - (5) - (6) + (3)$	39.08	18	2.17	

The computational procedures that have been discussed above can be specialized to the case of a $p \times q$ factorial experiment with repeated measures on factor *B*. This is done by setting $r = 1$ and dropping all terms involving factor *C*. The starting point for an unweighted-means analysis in this case is a summary table in which

$$AB'_{ij} = \frac{AB_{ij}}{n_i}.$$

A numerical example of this case is given in Table 7.8-3.

A summary of the computational steps in obtaining both the least-squares and the unweighted-means solutions appears in Table 7.8-4. Symbols associated with the least-squares solution appear at the left. The unweighted means analysis is summarized in Table 7.8-5. The two solutions are compared in Table 7.8-6.

Individual comparisons for the case of a $p \times q$ factorial experiment having repeated measures on factor *B* will be outlined in what follows. For the case of the least-squares solution,

$$\bar{A}_i = \frac{A_i}{n_i q} \quad \text{and} \quad \bar{B}_j = \frac{B_j}{N}.$$

Table 7.8-6 Comparison of Solutions

Source	Unweighted means	Least squares
<i>A</i>	23.17	22.94
<i>B</i>	346.73	365.06
<i>AB</i>	179.86	195.86



F ratios in tests on individual comparisons take the form

$$F = \frac{(\bar{A}_i - \bar{A}_{i'})^2}{MS_{\text{subj w. groups}} \left(\frac{1}{n_i q} + \frac{1}{n_{i'} q} \right)}$$

$$F = \frac{(\bar{B}_j - \bar{B}_{j'})^2}{MS_{B \times \text{subj w. groups}} (2/N)}$$

For the case of the unweighted-means solution,

$$\bar{A}_i = \frac{A'_i}{q} \quad \text{and} \quad \bar{B}_j = \frac{B'_j}{p}$$

F ratios in tests on individual comparisons take the form

$$F = \frac{(\bar{A}_i - \bar{A}_{i'})^2}{MS_{\text{subj w. groups}} \left(\frac{1}{n_i q} + \frac{1}{n_{i'} q} \right)}$$

$$F = \frac{(\bar{B}_j - \bar{B}_{j'})^2}{MS_{B \times \text{subj w. groups}} (2/\bar{n}_h p)}$$



CHAPTER 8

Factorial Experiments in Which Some of the Interactions Are Confounded

8.1 General Purpose

Precision of estimation requires that treatment effects be free of between-block variation. In this context, a block is a person, a group of people, a period of time, a source of experimental material, etc. Since the number of treatment combinations in a factorial experiment increases rapidly with either the number of factors or the number of levels of the factors, observing all treatment combinations within the same block often demands unworkably large block capacity. In this chapter, techniques will be considered for assigning a relatively small number of the possible treatment combinations to blocks in a way that permits within-block estimates of the most important sources of variation.

Small block size in this context is equivalent to homogeneity of the conditions under which the treatment effects are measured. There can be considerable variation among blocks; this source of variation does not affect the precision of the within-block information. When practical working conditions rule against having a complete replication within a single block and still assure homogeneity of the uncontrolled sources of error, then balanced incomplete-block designs provide the next best alternative.

Most of the plans to be developed use all the treatment combinations required for the complete factorial experiment. Within any one block only a fraction of all possible treatment combinations appear; the number of treatment combinations per block will be called the block size. The primary object of these plans is to control experimental error (1) by keeping the block size small and (2) by eliminating block differences from experimental error. The cost of this added control on experimental error will be the loss of some information on higher-order interactions. Differences between blocks will form a part of some interaction; hence components of such interactions will be confounded with the block effects. Whether or not the

sacrifice of information on components of higher-order interactions is worth the added control depends upon the magnitude of the block effects.

Admittedly there is some danger in using designs in which any effect is confounded. This is particularly true in exploratory studies. However, the potential advantage of these designs—increased precision with respect to effects of primary interest—provides the experimenter with a potent source of motivation for their use.

In this chapter the general principles underlying designs involving confounding of interactions will be considered; the analysis and applications of these designs will also be illustrated. In balanced designs some information will be available on all components of the interactions; in other designs some of the components of interactions will be completely confounded with between-block effects. The designs to be discussed in this chapter resemble those to be discussed in Chaps. 9 and 10. The designs in Chap. 10, which use the Latin-square principle, are actually special cases of the designs in this chapter. The incomplete-block designs presented in Chap. 9 are single-factor experiments rather than factorial experiments; however, quasi factors are introduced for purposes of the analysis.

Construction of the designs to be presented in this chapter depends upon techniques for analyzing interaction terms into parts. The first step is to divide the treatment combinations into sets which are balanced with respect to the main effects of the factors involved. For example, in a 2×2 factorial experiment the following sets are balanced with respect to the main effects:

Set I	Set II
ab_{11}	ab_{12}
ab_{22}	ab_{21}

Note that a_1 appears once and only once in each set. Similarly, a_2 , b_1 , and b_2 each appear once and only once in each set. Thus, there is balance with respect to the main effects of both factors A and B . Any difference between the sums for the sets is *not* a function of the main effects, but rather a function of the AB interaction. These results follow only if factors A and B are fixed; blocks are assumed to be random. Further, the underlying model is strictly additive with respect to block effects; i.e., no interactions with block effects appear in the model.

$$X_{ijkm} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + (\text{block})_k + \varepsilon_{m(jkm)}.$$

In obtaining the components of a three-factor interaction one divides all treatment combinations into balanced sets. In this case, however, the balance is with respect to all two-factor interactions as well as all main effects. For example, in a $2 \times 2 \times 2$ factorial experiment balanced sets with respect to the ABC interaction are:

Set I	Set II
abc_{111}	abc_{112}
abc_{122}	abc_{211}
abc_{212}	abc_{121}
abc_{221}	abc_{222}

Note that each level of each factor occurs an equal number of times within a set. Note also that all possible pairs of treatments occur once and only once with each set. Hence each set is balanced with respect to the A , B , and C main effects as well as AB , AC , and BC interactions. Therefore, the difference between the sum of all observations in set I and the sum of all observations in set II is a function of the ABC interaction.

Procedures for constructing balanced sets of treatment combinations are given in later sections. Comparisons between the sums over balanced sets define components of the interactions.

The sets of treatment combinations which are balanced with respect to main effects and lower-order interactions will be assigned to separate blocks. Hence the between-block differences, though free of main effects and lower-order interactions, will be confounded with components of the higher-order interactions. By replicating the experiment an appropriate number of times, different components of higher-order interactions can be used to form the sets of treatment combinations that make up the blocks. Hence some information on all components of the higher-order interactions will often be obtainable from the experiment as a whole.

8.2 Modular Arithmetic

The material to be presented in later sections is simplified through use of modular arithmetic. By definition an integer I modulus an integer m is the remainder obtained by dividing I by m . For example, the integer 18 to the modulus 5 is 3, since the remainder when 18 is divided by 5 is 3. This result is usually written

$$18 \pmod{5} = 3$$

and is read "18 modulo 5 is 3." Other examples are:

$$20 \pmod{5} = 0,$$

$$7 \pmod{5} = 2,$$

$$3 \pmod{5} = 3.$$

Alternatively, 18 is said to be congruent to 3, modulus 5; 20 is said to be congruent to 0, modulus 5. Thus, all integers are congruent to one of the integers 0, 1, 2, 3, or 4, modulus 5.

If the modulus 3 is used, all integers are congruent to 0, 1, or 2. For the modulus 3,

$$18 \pmod{3} = 0,$$

$$7 \pmod{3} = 1,$$

$$20 \pmod{3} = 2.$$

For purposes of the work in the following sections, the moduli will be limited to prime numbers, i.e., numbers divisible by no number smaller than it except unity. For example, 1, 2, 3, 5, 7, 11, etc., are prime numbers.

The operation of modular addition is shown in the following examples:

$$2 + 1 = 0 \pmod{3},$$

$$0 + 2 = 2 \pmod{3},$$

$$2 + 2 = 1 \pmod{3};$$

$$2 + 2 = 4 \pmod{5},$$

$$4 + 4 = 3 \pmod{5},$$

$$1 + 4 = 0 \pmod{5}.$$

To add two integers, one obtains the ordinary sum, and then one expresses this sum in terms of the modulus. For example, $4 + 4 = 8$; $8 \pmod{5} = 3$. Hence, $4 + 4 = 3 \pmod{5}$. Unless the modulus is understood from the context, it is written after the operation as \pmod{m} . The operation of multiplication is illustrated by the following examples:

$$2 \cdot 2 = 1 \pmod{3},$$

$$2 \cdot 0 = 0 \pmod{3};$$

$$4 \cdot 2 = 3 \pmod{5},$$

$$3 \cdot 3 = 4 \pmod{5}.$$

The product of two numbers is formed as in ordinary multiplication; then the product is expressed in terms of the modulus.

Algebraic equations may be solved in terms of a modular system. For example, by using the modulus 3, the equation

$$2x = 1 \pmod{3}$$

has the solution $x = 2$. To obtain the solution, both sides of this equation are multiplied by a number which will make the coefficient of x equal to unity, modulus 3. Thus,

$$2 \cdot 2x = 2 \pmod{3}.$$

Since $2 \cdot 2 = 1 \pmod{3}$, the last equation becomes $x = 2$. As another example, the equation

$$4x = 3 \pmod{5}$$

has the solution $x = 2$. To obtain this solution, both sides of the equation are multiplied by an integer that will make the coefficient of x equal to unity. If both sides of this equation are multiplied by 4,

$$4 \cdot 4x = 4 \cdot 3 \pmod{5}.$$

Expressing the respective products to the modulus 5, one has

$$16x = 12 \quad \text{or} \quad x = 2 \pmod{5}.$$

Equations of the form $ax_1 + bx_2 = c \pmod{m}$ may always be reduced to the form $x_1 + dx_2 = k \pmod{m}$. For example, the equation

$$2x_1 + x_2 = 1 \pmod{3}$$

becomes, after multiplying both sides by 2,

$$x_1 + 2x_2 = 2 \pmod{3}.$$

As another example,

$$2x_1 + 4x_2 = 2 \pmod{5}$$

becomes, after multiplying both sides by 3,

$$x_1 + 2x_2 = 1 \pmod{5}.$$

The equation $2x_1 + 4x_2 = 2 \pmod{5}$ and the equation $x_1 + 2x_2 = 1 \pmod{5}$ have the same roots. It may be verified that when $x_1 = 0$, $x_2 = 3$; hence one pair of roots for these equations is (0,3). Other roots are (1,0) and (2,2). To show that (2,2) is a root of the equation $x_1 + 2x_2 = 1 \pmod{5}$, substituting $x_1 = 2$ and $x_2 = 2$ in this equation yields

$$2 + 2(2) = 1 \pmod{5}.$$

The numerical value of the left-hand side of the equation is 6, which to the modulus 5 is 1.

8.3 Revised Notation for Factorial Experiments

The introduction of a modified notation for the treatment combinations in a factorial experiment will permit more convenient application of modular arithmetic. This revised notation is illustrated for the case of a $3 \times 3 \times 2$ factorial experiment in Table 8.3-1. The three levels of factor A are designated by the subscripts 0, 1, and 2. The treatment combination consisting of level a_1 , level b_0 , and level c_1 is designated by the symbol (101).

Table 8.3-1 Notation for a $3 \times 3 \times 2$ Factorial Experiment

	c_0			c_1		
	b_0	b_1	b_2	b_0	b_1	b_2
a_0	(000)	(010)	(020)	(001)	(011)	(021)
a_1	(100)	(110)	(120)	(101)	(111)	(121)
a_2	(200)	(210)	(220)	(201)	(211)	(221)

The digit in the first position indicates the level of factor A , the digit in the second position indicates the level of factor B , and the digit in the third position indicates the level of factor C . Thus, the symbol (ijk) represents the treatment combination abc_{ijk} .

To illustrate how modular arithmetic may be used to define sets of treatment combinations, consider a 3×3 factorial experiment. All the treatment combinations in this experiment may be expressed in the form (ij) by

suitable choice of i and j . Let x_1 stand for the digit in the first position of this symbol, and let x_2 stand for the digit in the second position of this symbol. The relation $x_1 + x_2 = 0 \pmod{3}$ is satisfied by the symbols (00), (12), and (21). To show, for example, that the symbol (12) satisfies this relation, substituting $x_1 = 1$ and $x_2 = 2$ yields $1 + 2 = 0 \pmod{3}$. Equivalently, the relation $x_1 + x_2 = 0$ is said to define or generate the set of treatment combinations (00), (12), (21). By similar reasoning, the relation

$x_1 + x_2 = 1 \pmod{3}$ defines the set (01), (10), (22);

and the relation

$x_1 + x_2 = 2 \pmod{3}$ defines the set (02), (11), (20).

Each of these sets is balanced with respect to main effects.

8.4 Method for Obtaining the Components of Interactions

The method of subdividing the degrees of freedom for interactions to be described in this section applies only to factorial experiments of the form $p \times p \times \cdots \times p$, where p is a prime number. Thus, this method applies to experiments of the form 2×2 , $2 \times 2 \times 2$, \dots ; 3×3 , $3 \times 3 \times 3$, \dots ; 5×5 , $5 \times 5 \times 5$, \dots ; etc. A 3×3 factorial experiment will be used to illustrate the method. The $A \times B$ interaction, which has four degrees of freedom, consists of an I component having two degrees of freedom and a J component having two degrees of freedom. The sums of squares for $A \times B$ may be partitioned as follows:

SS	df
$A \times B$	4
$AB(I)$	2
$AB(J)$	2

The symbol $AB(I)$ denotes the I component of the $A \times B$ interaction. Such components may have no meaning in terms of the levels of the factors; their purpose is merely to provide a convenient method for subdividing the degrees of freedom.

An alternative, but more convenient, notation for the components of the interaction uses the symbol AB for the J component and the symbol AB^2 for the I component. The rationale underlying this notation scheme will become clear when the computational procedure for these components is described. The nine treatment combinations in a 3×3 factorial experiment may be divided into three nonoverlapping sets by means of the following relations:

$$x_1 + x_2 = 0 \pmod{3}$$

(00)
(12)
(21)

$$x_1 + x_2 = 1 \pmod{3}$$

(01)
(10)
(22)

$$x_1 + x_2 = 2 \pmod{3}$$

(02)
(11)
(20)

The digits in the symbols for the treatments satisfy the relations under which the symbols appear. No treatment combination appears in more than one set; all treatment combinations of the 3×3 factorial experiment are included. Each of these sets is balanced with respect to the main effects of both factors (assuming that factors A and B are both fixed factors).

To illustrate what is meant by balance in this context, assume that the following linear model holds:

$$X_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \varepsilon_{ij},$$

where $\sum_i \alpha_i = 0$, $\sum_j \beta_j = 0$, and $\sum_i (\alpha\beta)_{ij} = \sum_j (\alpha\beta)_{ij} = 0$. When A and B are fixed factors, these restrictions on the parameters in the model follow from the definition of the effects. An observation made on a treatment combination on the left-hand side of the following equations estimates the sum of the parameters indicated on the right-hand side,

$$(00) = \mu + \alpha_0 + \beta_0 + (\alpha\beta)_{00} + \varepsilon_{00},$$

$$(12) = \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} + \varepsilon_{12},$$

$$(21) = \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} + \varepsilon_{21}.$$

The sum of the observations on this set of treatments will not contain effects associated with either of the main effects, since $\alpha_0 + \alpha_1 + \alpha_2 = 0$ and $\beta_0 + \beta_1 + \beta_2 = 0$. The sum of the observations in this set of treatments will, however, involve an interaction effect, since $(\alpha\beta)_{00} + (\alpha\beta)_{12} + (\alpha\beta)_{21} \neq 0$. Thus, the set (00), (12), (21), which is defined by the relation $x_1 + x_2 = 0 \pmod{3}$, is said to be balanced with respect to both main effects but not balanced with respect to the interaction effect. The sets defined by the relations $x_1 + x_2 = 1$ and $x_1 + x_2 = 2$ are also balanced with respect to both main effects but unbalanced with respect to the interaction effect.

Differences between the sums of the observations for each of these three sets define two of the four degrees of freedom of the $A \times B$ interaction. These two degrees of freedom are associated with what is called the $AB(J)$ component (or the AB component) of the $A \times B$ interaction. The numerical example in Table 8.4-1 illustrates the computation of this component of the interaction. The entries in the cells of part i of this table are assumed to be the sums over four observations; i.e., the data in part i represent a summary of a 3×3 factorial experiment in which there are four observations per cell.

The sum of squares for the over-all $A \times B$ interaction, which has four degrees of freedom, is computed in part ii. The sum of squares for the AB component, which has two degrees of freedom, is computed in part iii. The sum of all observations in the set of treatment combinations defined by the relation $x_1 + x_2 = 0$ is denoted by the symbol $(AB)_0$. Similarly the sum of all observations in the set defined by $x_1 + x_2 = 1$ is denoted by the

Table 8.4-1 Computation of Components of Interaction in 3×3 Factorial Experiment Having Four Observations per Cell

(i)

	b_0	b_1	b_2		
a_0	45	20	20	$85 = A_0$	$(n = 4)$
a_1	25	50	20	$95 = A_1$	
a_2	20	50	80	$150 = A_2$	
	90	120	120	$330 = G$	
	B_0	B_1	B_2		

(ii)

$$SS_{A \times B} = \frac{45^2 + 20^2 + \cdots + 50^2 + 80^2}{4} - \frac{85^2 + 95^2 + 150^2}{12} - \frac{90^2 + 120^2 + 120^2}{12} + \frac{330^2}{36}$$

$$= 633.33$$

(iii)

$x_1 + x_2 = 0$	$x_1 + x_2 = 1$	$x_1 + \bar{x}_2 = 2$
(00) = 45	(01) = 20	(02) = 20
(12) = 20	(10) = 25	(11) = 50
(21) = 50	(22) = 80	(20) = 20
$(AB)_0 = 115$	$(AB)_1 = 125$	$(AB)_2 = 90$

$$SS_{AB} = \frac{115^2 + 125^2 + 90^2}{12} - \frac{330^2}{36}$$

$$= 54.17$$

(iv)

$x_1 + 2x_2 = 0$	$x_1 + 2x_2 = 1$	$x_1 + 2x_2 = 2$
(00) = 45	(02) = 20	(01) = 20
(11) = 50	(10) = 25	(12) = 20
(22) = 80	(21) = 50	(20) = 20
$(AB^2)_0 = 175$	$(AB^2)_1 = 95$	$(AB^2)_2 = 60$

$$SS_{AB^2} = \frac{175^2 + 95^2 + 60^2}{12} - \frac{330^2}{36}$$

$$= 579.17$$

symbol $(AB)_1$. The sum of squares corresponding to the AB component is given by

$$SS_{AB} = \frac{(AB)_0^2 + (AB)_1^2 + (AB)_2^2}{3n} - \frac{G^2}{9n},$$

where n is the number of observations in each of the cell entries in part i. For the data in part i, $SS_{AB} = 54.17$.

The AB^2 component of the interaction is obtained from sets of treatment combinations defined by the following relations (modulus 3 is understood):

$x_1 + 2x_2 = 0$	$x_1 + 2x_2 = 1$	$x_1 + 2x_2 = 2$
(00)	(02)	(01)
(11)	(10)	(12)
(22)	(21)	(20)

These sets are also balanced with respect to the main effects. Hence differences among the sums of the observations in these three sets will define a component of the over-all interaction—in this case the AB^2 component. The numerical value of this component is computed in part iv of Table 8.4-1. The symbol $(AB^2)_0$ in this part denotes the sum of all observations in the set of treatments defined by the relation $x_1 + 2x_2 = 0$. The sum of squares for this component is given by

$$SS_{AB^2} = \frac{(AB^2)_0^2 + (AB^2)_1^2 + (AB^2)_2^2}{3n} - \frac{G^2}{9n}.$$

The numerical value for this sum of squares is 579.17.

The partition of the over-all interaction may be summarized as follows:

Source	SS	df
$A \times B$	633.33	4
AB	54.17	2
AB^2	579.17	2

Within rounding error $SS_{A \times B} = SS_{AB} + SS_{AB^2}$. In this case the AB^2 component of the interaction is considerably larger than the AB component. Inspection of part i of Table 8.4-1 indicates that the large totals are on the diagonal running from upper left to lower right. The treatment totals on this diagonal are in the same set for the AB^2 component but in different sets for the AB component. Hence the AB^2 component is large relative to the AB component.

Table 8.4-2 Simplified Computation of the Components of the $A \times B$ Interaction

	b_0	b_1	b_2	
a_0	45	20	20	
a_1	25	50	20	
$(AB)_2 = 90$	a_2 —20	50	80	—175 = $(AB^2)_0$
$(AB)_0 = 115$	a_0 —45	20	20	—95 = $(AB^2)_1$
$(AB)_1 = 125$	a_1 —25	50	20	—60 = $(AB^2)_2$

The computation of the components of the $A \times B$ interaction may be simplified by use of the procedure illustrated in Table 8.4-2. The cell entries in part i of this table are the same as those in part i of Table 8.4-1, with rows a_0 and a_1 appearing twice. Totals of the form $(AB)_i$ are obtained by summing the entries on diagonals running from upper right to lower left; totals

of the form $(AB^2)_i$ are obtained by summing entries on diagonals running from upper left to lower right. From these totals the components of the interaction are computed by the same formulas as those used in Table 8.4-1. The treatment combinations falling along these diagonals actually form the balanced sets defined by modular relations.

The use of modular arithmetic in defining balanced sets of treatment combinations is readily extended to 5×5 factorial experiments. In this latter type of experiment the 16 degrees of freedom for the $A \times B$ interaction may be partitioned as follows.

Source	df
$A \times B$	16
AB	4
AB^2	4
AB^3	4
AB^4	4

The balanced sets of treatment combinations from which the AB^3 component of the interaction is obtained are defined by the following relations, all with respect to the modulus 5:

$x_1 + 3x_2 = 0$	$x_1 + 3x_2 = 1$	$x_1 + 3x_2 = 2$	$x_1 + 3x_2 = 3$	$x_1 + 3x_2 = 4$
(00)	(02)	(04)	(01)	(03)
(13)	(10)	(12)	(14)	(11)
(21)	(23)	(20)	(22)	(24)
(34)	(31)	(33)	(30)	(32)
(42)	(44)	(41)	(43)	(40)

Each of these sets is balanced with respect to both main effects. The balanced sets defining the AB^4 component of the interaction are obtained from the relation $x_1 + 4x_2 = i$ (modulus 5), where $i = 0, \dots, 4$.

The computation of the components of the $A \times B$ interaction in a 5×5 factorial experiment may be carried out by an extension of the method used in Table 8.4-2. This extension is illustrated by the data in Table 8.4-3. In part i, the totals required for the computation of the AB component appear at the left, and the totals required for the AB^4 component appear at the right. In part ii, totals required for the AB^2 and the AB^3 components are obtained. The arrangement of the b 's in part ii is given by multiplying subscripts to b 's in part i by 2 and then reducing the resulting numbers to the modulus 5. By performing this operation, the sequence b_0, b_1, b_2, b_3, b_4 becomes the sequence b_0, b_2, b_4, b_1, b_3 ; the latter sequence appears in part ii. The AB^2 component of the interaction is obtained from the totals in the lower right of part ii. Assuming one observation per cell,

$$SS_{AB^2} = \frac{65^2 + 70^2 + 65^2 + 80^2 + 100^2}{5} - \frac{380^2}{25} = 174.00.$$

Table 8.4-3 Computation of Components of Interaction in a 5×5 Factorial Experiment

		b_0	b_1	b_2	b_3	b_4	
	a_0	5	10	15	20	25	
	a_1	10	15	20	25	30	
	a_2	15	15	15	15	20	
	a_3	20	15	10	10	5	
(i)	$(AB)_4 = 110$	a_4	30	20	5	5	$50 = (AB^4)_0$
	$(AB)_0 = 80$	a_0	5	10	15	20	$65 = (AB^4)_1$
	$(AB)_1 = 55$	a_1	10	15	20	25	$85 = (AB^4)_2$
	$(AB)_2 = 55$	a_2	15	15	15	20	$100 = (AB^4)_3$
	$(AB)_3 = 80$	a_3	20	15	10	10	$80 = (AB^4)_4$
							$\frac{380}{380}$

		b_0	b_2	b_4	b_1	b_3	
	a_0	5	15	25	10	20	
	a_1	10	20	30	15	25	
	a_2	15	15	20	15	15	
	a_3	20	10	5	15	10	
(ii)	$(AB^3)_4 = 95$	a_4	30	5	5	20	$65 = (AB^2)_0$
	$(AB^3)_0 = 55$	a_0	5	15	25	10	$70 = (AB^2)_1$
	$(AB^3)_1 = 60$	a_1	10	20	30	15	$65 = (AB^2)_2$
	$(AB^3)_2 = 90$	a_2	15	15	20	15	$80 = (AB^2)_3$
	$(AB^3)_3 = 80$	a_3	20	10	5	15	$100 = (AB^2)_4$
							$\frac{380}{380}$

Other components are obtained in an analogous manner. The numerical values of the components of the interaction for the data in Table 8.4-3 are as follows:

Source	SS	df
$A \times B$	1136.00	16
AB	414.00	4
AB^2	174.00	4
AB^3	254.00	4
AB^4	294.00	4

In a $3 \times 3 \times 3$ factorial experiment, the three-factor interaction may be partitioned into the following components:

Source	df
$A \times B \times C$	8
ABC	2
ABC^2	2
AB^2C	2
AB^2C^2	2

The sets of treatment combinations from which the AB^2C^2 component is computed are defined by the following relations:

$x_1 + 2x_2 + 2x_3 = 0 \pmod{3}$	$x_1 + 2x_2 + 2x_3 = 1 \pmod{3}$	$x_1 + 2x_2 + 2x_3 = 2 \pmod{3}$
(000)	(002)	(001)
(012)	(011)	(010)
(021)	(020)	(022)
(101)	(100)	(102)
(110)	(112)	(111)
(122)	(121)	(120)
(202)	(201)	(200)
(211)	(210)	(212)
(220)	(222)	(221)

The sum of the observations within each of these sets is balanced with respect to all the main effects and all the two-factor interactions. There is no balance for the three-factor interaction. Hence differences between sums over the sets form one of the components of the three-factor interaction, in this case the AB^2C^2 component. Treatment combinations from which the AB^2C component is computed are defined by relations of the form $x_1 + 2x_2 + x_3 = i \pmod{3}$.

The computation of the components of the three-factor interaction is illustrated in Table 8.4-4. In part i summary data for observations made at level c_0 are given. Summary data for levels c_1 and c_2 appear in parts ii and iii, respectively. Totals corresponding to the AB component for level c_0

appear at the left in part i. These totals also appear in part iv under the column headed c_0 . The totals at the left of parts ii and iii make up the columns c_1 and c_2 in part iv. The ABC component of the three-factor interaction is computed from totals at the left of part iv; the ABC^2 component is obtained from the totals at the right of part iv. For example, if there are

Table 8.4-4 Computation of the Components of a Three-Factor Interaction in a $3 \times 3 \times 3$ Factorial Experiment

(i)

		c_0			
		b_0	b_1	b_2	
	a_0	126	63	71	
	a_1	173	106	170	
$(AB)_2 = 263$	a_2	86	193	164	$-396 = (AB^2)_0$
$(AB)_0 = 489$	a_0	126	63	71	$-437 = (AB^2)_1$
$(AB)_1 = 400$	a_1	173	106	170	$-319 = (AB^2)_2$

(ii)

		c_1			
		b_0	b_1	b_2	
	a_0	127	158	106	
	a_1	80	109	123	
$(AB)_2 = 307$	a_2	92	139	104	$-340 = (AB^2)_0$
$(AB)_0 = 389$	a_0	127	158	106	$-325 = (AB^2)_1$
$(AB)_1 = 342$	a_1	80	109	123	$-373 = (AB^2)_2$

(iii)

		c_2			
		b_0	b_1	b_2	
	a_0	103	80	54	
	a_1	123	52	109	
$(AB)_2 = 198$	a_2	92	101	60	$-215 = (AB^2)_0$
$(AB)_0 = 313$	a_0	103	80	54	$-278 = (AB^2)_1$
$(AB)_1 = 263$	a_1	123	52	109	$-281 = (AB^2)_2$

Table 8.4-4 (cont.)

		c_0	c_1	c_2	
	$(AB)_0$	489	389	313	
	$(AB)_1$	400	342	263	
(iv)	$(ABC)_2 = 918$	$(AB)_2 = 263$	307	198	$1029 = (ABC^2)_0$
	$(ABC)_0 = 1059$	$(AB)_0 = 489$	389	313	$1020 = (ABC^2)_1$
	$(ABC)_1 = 987$	$(AB)_1 = 400$	342	263	$915 = (ABC^2)_2$
		<u>2964</u>			<u>2964</u>

		c_0	c_1	c_2	
	$(AB^2)_0$	396	340	215	
	$(AB^2)_1$	437	325	278	
(v)	$(AB^2C)_2 = 859$	$(AB^2)_2 = 319$	373	281	$1002 = (AB^2C^2)_0$
	$(AB^2C)_0 = 1047$	$(AB^2)_0 = 396$	340	215	$1025 = (AB^2C^2)_1$
	$(AB^2C)_1 = 1058$	$(AB^2)_1 = 437$	325	278	$937 = (AB^2C^2)_2$

r observations in each cell of a $3 \times 3 \times 3$ factorial experiment, the ABC component is given by

$$SS_{ABC} = \frac{918^2 + 1059^2 + 987^2}{9r} - \frac{2964^2}{27r}.$$

In part v of the table, the entries under column c_0 are obtained from the totals at the right of part i. The entries in columns c_1 and c_2 are obtained, respectively, from the totals to the right of parts ii and iii. Totals at the left of part v are used to obtain the AB^2C component of the three-factor interaction; totals at the right are used to obtain the AB^2C^2 component.

The treatment combinations that are summed to obtain the total $(AB^2C^2)_0$ in part v may be shown to be those belonging to set $x_1 + 2x_2 + 2x_3 = 0$. Similarly the total $(AB^2C^2)_1$ is the sum of all observations on the treatment set $x_1 + 2x_2 + 2x_3 = 1$. The scheme presented in Table 8.4-4 provides a convenient method for obtaining the sum over balanced sets. In general the symbol $(AB^iC^j)_k$ represents the sum of all treatment combinations satisfying the relation

$$x_1 + ix_2 + jx_3 = k \pmod{3}.$$

Rather than using the scheme presented in Table 8.4-4, it is sometimes more convenient to form the sets of treatment combinations by means of their defining relations and obtain the sums directly from the sets.

In the notation system used by Yates as well as Cochran and Cox, the following components of the three-factor interaction in a $3 \times 3 \times 3$ factorial experiment are equivalent:

Modular notation	Yates notation
ABC	$ABC(Z)$
ABC^2	$ABC(Y)$
AB^2C	$ABC(X)$
AB^2C^2	$ABC(W)$

The modular-notation system is the more convenient and lends itself to generalization beyond the $3 \times 3 \times 3$ factorial experiment. For example, in a $5 \times 5 \times 5$ factorial experiment, the three-factor interaction may be partitioned into the following parts:

Source	df
$A \times B \times C$	64
ABC	4
ABC^2	4
ABC^3	4
ABC^4	4
AB^2C	4
...	...
AB^4C^4	4

There are 16 components in all, each having four degrees of freedom. The balanced sets used to obtain the AB^4C^4 component are defined by the relations $x_1 + 4x_2 + 4x_3 = 0, 1, 2, 3, 4 \pmod{5}$. For example, the sum $(AB^4C^4)_2$ is obtained from the treatment combinations belonging to the set defined by the relation $x_1 + 4x_2 + 4x_3 = 2$; there will be 25 treatment combinations in this set.

The notation system for the components of the interaction is not unique unless the convention is adopted that the exponent of the first letter is always unity. For example, in a 3×3 factorial experiment A^2B and AB^2 define the same component, since

$$2x_1 + x_2 = i \pmod{3}$$

defines the same set of treatment combinations as

$$2(2)x_1 + 2x_2 = 2i \pmod{3},$$

$$x_1 + 2x_2 = 2i \pmod{3}.$$

Multiplying both sides of a defining relation by a constant is algebraically equivalent to raising the symbol for the corresponding component to a power equal to that constant. In this case the constant is 2. Hence

$$(A^2B)^2 = A^4B^2 = AB^2,$$

upon expressing the exponents modulo 3.

As another example, in a 5×5 factorial experiment A^3B^2 is equivalent to AB^4 . To show this,

$$3x_1 + 2x_2 = i \pmod{5}.$$

Multiplying both sides of this equation by 2,

$$6x_1 + 4x_2 = 2i \pmod{5},$$

$$x_1 + 4x_2 = 2i \pmod{5}.$$

Alternatively,

$$A^3B^2 = (A^3B^2)^2 = A^6B^4 = AB^4.$$

8.5 Designs for $2 \times 2 \times 2$ Factorial Experiments in Blocks of Size 4

The eight treatment combinations in a $2 \times 2 \times 2$ factorial experiment may be divided into two balanced sets in such a way that all main effects and two-factor interactions are balanced within each set. The plan in Table 8.5-1 provides such sets. The treatment combinations in block 1 satisfy the

Table 8.5-1 $2 \times 2 \times 2$ Factorial Experiment in Blocks of Size 4

Block 1	Block 2
(000)	(001)
(011)	(010)
(101)	(100)
(110)	(111)

relation $x_1 + x_2 + x_3 = 0 \pmod{2}$, and those in block 2 satisfy the relation $x_1 + x_2 + x_3 = 1 \pmod{2}$. For r replications of this experiment the analysis takes the form given in Table 8.5-2. It is assumed that the blocks are random but that factors A , B , and C are fixed. It is also assumed that the treatment by block interactions is negligible relative to other effects in the experiment.

The tests on main effects and two-factor interactions use the within-block residual in the denominator of F ratios. The test on the three-factor interaction uses the between-block residual as the denominator. This latter test will generally be considerably less sensitive than tests which are based upon within-block information. The within-block residual in this design is the pooled interaction of the replications with the main effects and with the two-factor interactions.

The plan in Table 8.5-1 does not provide any within-block information on the three-factor interaction. By way of contrast the plan in Table 8.5-3 provides within-block information on the three-factor interaction as well as the two-factor interactions. A minimum of eight blocks is required for this latter design. In the first replication, AB is confounded with differences between blocks 1 and 2. However, within-block information on AB is

available from all other replications. The treatment combinations in block 1 satisfy the relation $x_1 + x_2 = 0 \pmod{2}$; the treatment combinations appearing in block 2 satisfy the relation $x_1 + x_2 = 1 \pmod{2}$. Since

Table 8.5-2 Analysis of $2 \times 2 \times 2$ Factorial Experiment in Blocks of Size 4

Source of variation	df	E(MS)
Between blocks	$2r - 1$	
Replications	$r - 1$	
Blocks within reps	r	
ABC	1	$\sigma_e^2 + 4\sigma_{\text{blocks}}^2 + r\sigma_{\alpha\beta\gamma}^2$
Residual (between blocks)	$r - 1$	$\sigma_e^2 + 4\sigma_{\text{blocks}}^2$
Within blocks	$6r$	
A	1	$\sigma_e^2 + 4r\sigma_\alpha^2$
B	1	$\sigma_e^2 + 4r\sigma_\beta^2$
C	1	$\sigma_e^2 + 4r\sigma_\gamma^2$
AB	1	$\sigma_e^2 + 2r\sigma_{\alpha\beta}^2$
AC	1	$\sigma_e^2 + 2r\sigma_{\alpha\gamma}^2$
BC	1	$\sigma_e^2 + 2r\sigma_{\beta\gamma}^2$
Residual (within block)	$6r - 6$	σ_e^2

within-block information on AB is available from six of the eight blocks, the relative within-block information on the AB interaction is $\frac{6}{8} = \frac{3}{4}$.

To show that block 3, for example, provides within-block information on AB , two of the four treatment combinations in this block satisfy the relation

Table 8.5-3 Balanced Design with Partial Within-block Information on All Interactions

Rep 1		Rep 2		Rep 3		Rep 4	
Block 1	Block 2	Block 3	Block 4	Block 5	Block 6	Block 7	Block 8
(000)	(010)	(000)	(001)	(000)	(001)	(000)	(001)
(001)	(011)	(010)	(011)	(011)	(010)	(011)	(010)
(110)	(100)	(101)	(100)	(100)	(101)	(101)	(100)
(111)	(101)	(111)	(110)	(111)	(110)	(110)	(111)
AB		AC		BC		ABC	

$x_1 + x_2 = 0$ and hence belong to what will be called the set J_0 , that is, (000) and (110). The other two treatment combinations in block 3 belong to the set J_1 , that is, satisfy the relation $x_1 + x_2 = 1$. The difference between the totals for the J_0 set and the J_1 set provides information on the

AB component, which is free of block effects. In addition to block 3, each of blocks 4 through 8 contains two treatment combinations belonging to set J_0 and two belonging to set J_1 .

The AC interaction is completely confounded with the difference between blocks 3 and 4, but within-block information on AC is available from all other blocks. Hence the relative within-block information on AC is $\frac{6}{8}$ or $\frac{3}{4}$. In block 3, the treatment combinations satisfy the relation $x_1 + x_3 = 0 \pmod{2}$; in block 4 the treatment combinations satisfy the relation $x_1 + x_3 = 1 \pmod{2}$. Blocks 7 and 8 in Table 8.5-3 are identical to blocks 1 and 2 of the design in Table 8.5-1. In the latter design, no within-block

Table 8.5-4 Analysis of Balanced Design

Source of variation	df	E(MS)
Between blocks	7	
Replications	3	
Blocks within reps	4	
AB	1	
AC	1	
BC	1	
ABC	1	
Within blocks	24	
A	1	$\sigma_\varepsilon^2 + 16\sigma_\alpha^2$
B	1	$\sigma_\varepsilon^2 + 16\sigma_\beta^2$
C	1	$\sigma_\varepsilon^2 + 16\sigma_\gamma^2$
$(AB)'$	1	$\sigma_\varepsilon^2 + (\frac{3}{4})8\sigma_{\alpha\beta}^2$
$(AC)'$	1	$\sigma_\varepsilon^2 + (\frac{3}{4})8\sigma_{\alpha\gamma}^2$
$(BC)'$	1	$\sigma_\varepsilon^2 + (\frac{3}{4})8\sigma_{\beta\gamma}^2$
$(ABC)'$	1	$\sigma_\varepsilon^2 + (\frac{3}{4})4\sigma_{\alpha\beta\gamma}^2$
Residual	17	σ_ε^2

information is available on the ABC interaction; the design in Table 8.5-3 provides $\frac{3}{4}$ relative within-block information on ABC , such information being available from blocks 1 through 6.

The analysis of the design in Table 8.5-3 takes the form given in Table 8.5-4. The symbol $(AB)'$ indicates that only partial within-block information is available for AB . In the computation of $(AB)'$, only information from blocks 3 through 8 is used. The AB comparison given by the difference between block 1 and block 2 is confounded by differences between these blocks; this comparison represents one of the four degrees of freedom for blocks within replications. In computing $(AC)'$, only information from blocks 1, 2, and 5 through 8 is used. Similarly, $(ABC)'$ is based upon information from blocks 1 through 6. The difference between the total for

block 7 and the total for block 8 gives rise to the between-block component of ABC .

Each of the main effects is determined from data obtained from all four replications; the within-block information on the interactions is based upon data obtained in each case from three of the four replications. Hence there are three *effective* replications for the interactions, but four *effective* replications for the main effects. The degrees of freedom for the within-block residual are made up of the following parts:

Residual	17
$A \times \text{reps}$	3
$B \times \text{reps}$	3
$C \times \text{reps}$	3
$(AB)' \times \text{reps}$	2
$(AC)' \times \text{reps}$	2
$(BC)' \times \text{reps}$	2
$(ABC)' \times \text{reps}$	2

Since each of the interactions has only three effective replications, their interaction with replications has only two degrees of freedom.

The expected values for the mean squares given in Table 8.5-4 assume that there is one observation per cell. Suppose that each block represents a group of n subjects and that the groups are assigned at random to one of the eight blocks. Also suppose each subject within the groups is observed under each of the four treatment combinations in the block to which the group is assigned. Further assume that terms of the form $\sigma_{\alpha\pi}^2$ are equal to zero. Under these assumptions the analysis takes the form given in Table 8.5-5. In this analysis the individual subject is considered to be the experimental unit. Tests on main effects and interactions use the residual mean square in the denominator of the F ratio. There is usually no interest in a test on differences between blocks or the components of such differences. In some cases between-block estimates of the interactions may be combined with the corresponding within-block estimates. Such pooled estimates are given by a weighted sum of the between-subject and within-subject components, the weights being the respective reciprocals of the between- and within-person residuals. F tests on such combined estimates require a weighted pooling of the respective residuals.

The cell totals for the design in Table 8.5-3 may be designated by the following notation:

	b_0		b_1	
	c_0	c_1	c_0	c_1
a_0	$X_{000.}$	$X_{001.}$	$X_{010.}$	$X_{011.}$
a_1	$X_{100.}$	$X_{101.}$	$X_{110.}$	$X_{111.}$

Table 8.5-5 Analysis of Balanced Design with Repeated Measures

Source of variation	df	E(MS)
Between subjects	$8n - 1$	
Groups	7	
Reps	3	
<i>AB</i>	1	$\sigma_e^2 + 4\sigma_{\text{people}}^2 + 2n\sigma_{\alpha\beta}^2$
<i>AC</i>	1	$\sigma_e^2 + 4\sigma_{\text{people}}^2 + 2n\sigma_{\alpha\gamma}^2$
<i>BC</i>	1	$\sigma_e^2 + 4\sigma_{\text{people}}^2 + 2n\sigma_{\beta\gamma}^2$
<i>ABC</i>	1	$\sigma_e^2 + 4\sigma_{\text{people}}^2 + n\sigma_{\alpha\beta\gamma}^2$
Subjects within groups	$8(n - 1)$	$\sigma_e^2 + 4\sigma_{\text{people}}^2$
Within subjects	$24n$	
<i>A</i>	1	$\sigma_e^2 + 16n\sigma_{\alpha}^2$
<i>B</i>	1	$\sigma_e^2 + 16n\sigma_{\beta}^2$
<i>C</i>	1	$\sigma_e^2 + 16n\sigma_{\gamma}^2$
<i>(AB)'</i>	1	$\sigma_e^2 + (\frac{3}{4})8n\sigma_{\alpha\beta}^2$
<i>(AC)'</i>	1	$\sigma_e^2 + (\frac{3}{4})8n\sigma_{\alpha\gamma}^2$
<i>(BC)'</i>	1	$\sigma_e^2 + (\frac{3}{4})8n\sigma_{\beta\gamma}^2$
<i>(ABC)'</i>	1	$\sigma_e^2 + (\frac{3}{4})4n\sigma_{\alpha\beta\gamma}^2$
Residual	$24n - 7$	σ_e^2

Each of these totals will be based upon four observations. Since all the treatment combinations do not appear in all the blocks, means based upon these totals will not be free of block effects. If, for example, the cell total for treatment combination (101) were free of block effects, it would estimate

$$\begin{aligned}
 X_{101} &\doteq 4\mu + 4\alpha_1 + 4\beta_0 + 4\gamma_1 + 4(\alpha\beta)_{10} + 4(\alpha\gamma)_{11} \\
 &\quad + 4(\beta\gamma)_{01} + 4(\alpha\beta\gamma)_{101} \\
 &\quad + (\text{sum of all block effects}).
 \end{aligned}$$

In this last expression, the sum of all block effects would be a constant (under the assumptions made); hence differences between any two X 's would be free of block effects. It is readily verified that the following quantity will estimate the sum of the parameters on the right-hand side of the last expression:

$$X'_{101} = \frac{A_1 + B_0 + C_1}{4} + \frac{(AB)'_1 + (AC)'_0 + (BC)'_1 + (ABC)'_0}{3} - \frac{3G}{4},$$

where $(AB)'_1$ = total of all treatment combinations which satisfy the relation $x_1 + x_2 = 1 \pmod{2}$ and appear in blocks 3 through 8; similarly $(AC)'_0$ = total of all treatment combinations which satisfy the relation $x_1 + x_3 = 0 \pmod{2}$ and appear in blocks 1, 2, 4 through 8. The quantity X'_{101} is called

an adjusted cell total. The adjusted cell mean for treatment combination (101) is

$$\bar{X}'_{101} = \frac{X'_{101.}}{4}.$$

For the general case, an adjusted cell total for the design in Table 8.5-3 has the form

$$X'_{ijk.} = \frac{A_i + B_j + C_k}{4} + \frac{(AB)'_{i+j} + (AC)'_{i+k} + (BC)'_{j+k}}{3} + \frac{(ABC)'_{i+j+k}}{3} - \frac{3G}{4},$$

where $(AB)'_{i+j}$ = sum of all treatment combinations which satisfy the relation $x_1 + x_2 = (i + j) \pmod{2}$ and which appear in blocks providing within-block information on AB . Similarly $(ABC)'_{i+j+k}$ = sum of all treatment combinations which satisfy the relation $x_1 + x_2 + x_3 = (i + j + k) \pmod{2}$ and which appear in blocks providing within-block information on ABC .

Adjusted totals having the form $X'_{ij..}$ are obtained from the relation

$$X'_{ij..} = X'_{ij0.} + X'_{ij1.}.$$

Similarly,

$$X'_{i.k.} = X'_{i0k.} + X'_{i1k.}.$$

Individual comparisons among the treatment combinations use the adjusted cell totals. The effective number of observations in the adjusted cell total $X'_{ijk.}$ is actually somewhat less than four since there is only $\frac{3}{4}$ relative within-block information on the interaction effects. For most practical purposes a comparison between two adjusted cell totals is given by

$$F = \frac{(X'_{ijk.} - X'_{pqr.})^2}{8MS_{\text{res}}}.$$

$$F = \frac{(X'_{ij.} - X'_{pq.})^2}{16MS_{\text{res}}}.$$

A numerical example of the design given in Table 8.5-1 appears in Sec. 8.7. Computational procedures are simplified by means of special procedures for 2^k factorial experiments given in the next section.

8.6 Simplified Computational Procedures for 2^k Factorial Experiments

Computational procedures for 2×2 , $2 \times 2 \times 2$, $2 \times 2 \times 2 \times 2$, etc., factorial experiments may be simplified by use of the device employed in forming balanced sets. This simplified procedure was presented in a different form in Sec. 6.12. The simplified computational procedures will be illustrated by the numerical example given in Table 8.6-1. Cell totals in part i are the sum of five observations. The sums of squares obtained

by the procedures to be outlined are the *unadjusted* sums of squares if the experiment involves any confounding.

Table 8.6-1 $2 \times 2 \times 2 \times 2$ Factorial Experiment with Five Observations per Cell

(i)		b_0		b_1		Total
		c_0	c_1	c_0	c_1	
	a_0	5	10	15	15	45
	a_1	10	20	20	5	55
		15	30	35	20	100

	Treat comb: (000) (001) (010) (011) (100) (101) (110) (111)									Comparison
	Cell total:	5	10	5	15	10	20	20	5	
G		+	+	+	+	+	+	+	+	100
A		-	-	-	-	+	+	+	+	10
B		-	-	+	+	-	-	+	+	10
(ii) C		-	+	-	+	-	+	-	+	0
AB		+	+	-	-	-	-	+	+	-20
AC		+	-	+	-	-	+	-	+	-10
BC		+	-	-	+	+	-	-	+	-30
ABC		-	+	+	-	+	-	-	+	-20

$$SS_A = \frac{(10)^2}{5(8)} = 2.50$$

$$SS_{AC} = \frac{(-10)^2}{5(8)} = 2.50$$

$$(iii) \quad SS_B = \frac{(10)^2}{5(8)} = 2.50$$

$$SS_{BC} = \frac{(-30)^2}{5(8)} = 22.50$$

$$SS_C = 0$$

$$SS_{ABC} = \frac{(-20)^2}{5(8)} = 10.00$$

$$SS_{AB} = \frac{(-20)^2}{5(8)} = 10.00$$

The patterns of the algebraic signs in part ii are determined as follows (the x 's refer to digits in the treatment combinations at the top of each column; all addition is modulo 2):

$G:$	All positive		
$A:$	- if x_1	$= 0,$	+ if x_1 $= 1$
$B:$	- if x_2	$= 0,$	+ if x_2 $= 1$
$C:$	- if x_3	$= 0,$	+ if x_3 $= 1$
$AB:$	+ if $x_1 + x_2$	$= 0,$	- if $x_1 + x_2$ $= 1$
$AC:$	+ if $x_1 + x_3$	$= 0,$	- if $x_1 + x_3$ $= 1$
$BC:$	+ if $x_2 + x_3$	$= 0,$	- if $x_2 + x_3$ $= 1$
$ABC:$	- if $x_1 + x_2 + x_3$	$= 0,$	+ if $x_1 + x_2 + x_3$ $= 1$

In general if the number of factors which are interacting is even, then the zero modular sum receives the positive sign; if the number of factors interacting is odd, then the zero modular sum receives a negative sign. In this context main effects are classified with the interaction of an odd number of factors.

The entry at the right of each row in part ii is a weighted sum of the cell totals at the top of each column, the weights being ± 1 as determined by the pattern of signs in each row. For example, the entry at the right of row *A* is given by

$$-5 - 10 - 15 - 15 + 10 + 20 + 20 + 5 = 10.$$

As another example, the entry at the right of row *ABC* is given by

$$-5 + 10 + 15 - 15 + 10 - 20 - 20 + 5 = -20.$$

The pattern of the signs in the last expression is that in row *ABC*. The weighted sum of the cell totals for row *ABC* actually corresponds to the difference $(ABC)_1 - (ABC)_0$, where $(ABC)_1$ represents the sum of all observations on treatment combinations which satisfy the relation

$$x_1 + x_2 + x_3 = 1 \pmod{2}.$$

Alternatively, the totals at the right of the rows represent the comparisons corresponding to the effects at the left of the rows. The general form of a comparison (assuming the same number of observations in each cell total) is

$$\text{Comparison} = c_1 T_1 + c_2 T_2 + \cdots + c_k T_k, \quad \sum c = 0,$$

where the T 's represent cell totals. The sum of squares corresponding to a comparison is given by

$$SS_{\text{comparison}} = \frac{(\text{comparison})^2}{n \sum c^2},$$

where n is the number of observations summed to obtain the cell totals. The c 's in part ii are either $+1$ or -1 , depending upon the sign; for each of the rows $\sum c^2 = 8$. The sums of squares for comparisons corresponding to main effects and interactions are computed in part iii. In this example, n is assumed to be equal to 5.

The extension of these computational procedures to any factorial experiment of the form 2^k is direct. Once the pattern of signs corresponding to a row in part ii is determined, the weighted sum of the cell totals corresponding to a row gives the numerical value of the comparison. For any comparison determined in this manner, $\sum c^2 = 2^k$.

In 2^k factorial experiments a specialized notation is frequently used to designate the treatment combinations. This specialized notation for a $2 \times 2 \times 2$ factorial experiment is as follows:

	b_0		b_1	
	c_0	c_1	c_0	c_1
a_0	(1)	c	b	bc
a_1	a	ac	ab	abc

The relationship between the notation system that has been used so far in this chapter and the specialized notation system is as follows:

Treatment combination ...	(000)	(010)	(001)	(011)	(100)	(110)	(101)	(111)
Specialized notation	(1)	b	c	bc	a	ab	ac	abc

In the specialized notation system, the symbol for a treatment combination contains only those letters for which the factor is at level 1. Conversely, if a factor is at level 0 in a treatment combination, then the letter corresponding to that factor does not appear in the symbol.

In terms of the latter notation system sign patterns for the comparisons may be determined by a relatively simple rule. In the sign pattern for an interaction of the form XY , a symbol receives a positive sign if it contains both of the letters X and Y or neither of these letters; a symbol receives a negative sign otherwise. In the sign pattern for an interaction of the form XYZ , a symbol receives a positive sign if it contains all three of the letters X , Y , and Z or just one of the three letters; a symbol receives a negative sign otherwise. For the general case, in the interaction of an *even* number of factors, a symbol receives a positive sign if it contains an even number of the interacting factors (zero being considered an even number); otherwise it receives a negative sign. In the interaction of an *odd* number of factors, a symbol receives a positive sign if it contains an odd number of the interacting factors; otherwise the symbol receives a minus sign. A main effect may be considered included under the latter case. The following example illustrates the general case:

	Symbols:	(1)	b	c	bc	a	ab	ac	abc
A		—	—	—	—	+	+	+	+
C		—	—	+	+	—	—	+	+
AC		+	+	—	—	—	—	+	+

For the case of a 2^4 factorial experiment, the sign patterns for some of the comparisons are illustrated below:

	(1)	<i>a</i>	<i>b</i>	<i>ab</i>	<i>c</i>	<i>ac</i>	<i>bc</i>	<i>abc</i>	<i>d</i>	<i>ad</i>	<i>bd</i>	<i>abd</i>	<i>cd</i>	<i>acd</i>	<i>bcd</i>	<i>abcd</i>
<i>A</i>	—	+	—	+	—	+	—	+	—	+	—	+	—	+	—	+
<i>AB</i>	+	—	—	+	+	—	—	+	+	—	—	+	+	—	—	+
<i>ABC</i>	—	+	+	—	+	—	—	+	—	+	+	—	+	—	—	+
<i>ABCD</i>	+	—	—	+	—	+	+	—	—	+	+	—	+	—	—	+

Given the pattern of signs corresponding to a comparison, the sum of squares for that comparison is readily obtained from the treatment totals.

A systematic computing scheme for obtaining the comparisons corresponding to main effects and interactions is illustrated in Table 8.6-2. This

Table 8.6-2 Simplified Computational Procedures

Effect:						
	Treatment combination	Total	(1)	(2)	(3)	Comparison
(i)	(1)	5	15	50	100	<i>G</i>
	<i>a</i>	10	35	50	10	<i>A</i>
	<i>b</i>	15	30	10	10	<i>B</i>
	<i>ab</i>	20	20	0	-20	<i>AB</i>
	<i>c</i>	10	5	20	0	<i>C</i>
	<i>ac</i>	20	5	-10	-10	<i>AC</i>
	<i>bc</i>	15	10	0	-30	<i>BC</i>
	<i>abc</i>	5	-10	-20	-20	<i>ABC</i>
(ii)	Upper half	100	110	100	
	Lower half	10	-10	-60	
	Odds	45	60	80		
	Evens	55	50	20		

scheme is quite useful when the number of factors is large. The basic data in Table 8.6-2 are obtained from part i of Table 8.6-1. The treatment combinations must be arranged in the order given in this table. (If a fourth factor were present, the order for the treatment combinations having this factor at level 1 would be *d*, *ad*, *bd*, *abd*, *cd*, *acd*, *bcd*, *abcd*.)

The entries in the total column are the cell totals obtained from part i of Table 8.6-1. The entries in the upper half of column 1 are the sums of successive pairs of the entries in the total column, i.e., $5 + 10 = 15$, $15 + 20 = 35$, $10 + 20 = 30$, and $15 + 5 = 20$. The entries in the lower half of column 1 are differences between successive pairs in the total column, i.e., $10 - 5 = 5$, $20 - 15 = 5$, $20 - 10 = 10$ and $5 - 15 = -10$.

The entries in the upper half of column 2 are the sums of successive pairs of entries in column 1, that is, $15 + 35 = 50$, $30 + 20 = 50$. The entries in the lower half of column 2 are the differences between successive pairs of entries in column 1, that is, $35 - 15 = 20$, $20 - 30 = -10$.

The entries in column 3 are obtained from the entries in column 2 by a procedure analogous to that by which the entries in column 2 were obtained from column 1. This procedure is continued until column k is reached, where k is the number of factors. In this example $k = 3$; hence the procedure is terminated at column 3. The entries in column 3 give the numerical values of the comparisons corresponding to the treatment effects to the left of each row. These values are identical to those obtained in part ii of Table 8.6-1.

Checks on the numerical work are given in part ii of Table 8.6-2. The respective sums for the upper and lower halves for each of the columns are obtained. One then obtains the sum of every other entry in each column, starting with the first entry, i.e., the sum of odd-numbered entries. One also obtains the sum of the even-numbered entries in each of the columns. These sums are shown in part ii. The sum of the upper half of column 1 is checked by the sum of the odds and evens under the total column; the sum of the lower half of column 1 is checked by the difference between the evens and odds in the total column. The sum of the upper half of column 2 is checked by the sum of the odds and the evens under column 1; the sum of the lower half of column 2 is checked by the difference between evens and odds for column 1. Analogous checks are made for column 3.

8.7 Numerical Example of $2 \times 2 \times 2$ Factorial Experiment in Blocks of Size 4

The purpose of this experiment was to evaluate the effects of various treatment combinations upon the progress of mental patients in specified diagnostic categories. The treatments and their levels are given in Table 8.7-1. It was desired to have 10 patients in each cell of a factorial experiment, necessitating a total of 80 patients for the experiment. However, no more than 20 patients meeting specifications were available from a single

Table 8.7-1 Definition of Factors

Factor	Level	Definition
Drug <i>A</i>	a_0	No drug <i>A</i>
	a_1	Drug <i>A</i> administered
Drug <i>B</i>	b_0	No drug <i>B</i>
	b_1	Drug <i>B</i> administered
Psychotherapy	c_0	No psychotherapy
	c_1	Psychotherapy administered

hospital, but four hospitals were available for the study. It was anticipated that there would be large differences among hospitals. In order to prevent such differences from confounding main effects and two-factor interactions, the design in Table 8.7-2 was used.

Table 8.7-2 Design and Data for Numerical Example

Hospital 1	Hospital 2	Hospital 3	Hospital 4
<i>a</i> 6	(1) 2	<i>a</i> 14	(1) 3
<i>b</i> 10	<i>ab</i> 4	<i>b</i> 15	<i>ab</i> 6
<i>c</i> 6	<i>ac</i> 15	<i>c</i> 9	<i>ac</i> 25
<i>abc</i> 8	<i>bc</i> 18	<i>abc</i> 12	<i>bc</i> 22
30	39	50	56

$$\Sigma \Sigma X^2 = 2605$$

In the construction of this design, the relation $x_1 + x_2 + x_3 = i \pmod{2}$ was used to divide the eight treatment combinations into blocks of size 4. By using the specialized notation, this procedure is equivalent to assigning symbols containing an even number of letters to one set of blocks and symbols containing an odd number of letters to a second set of blocks. (In terms of the design in Table 8.7-2, symbols having an even number of letters appear in blocks 2 and 4.) Hospitals were assigned at random to blocks. Hospitals 1 and 2 make up one replication; hospitals 3 and 4 make up the second replication. Within each replication, the *ABC* interaction is completely confounded with between-hospital differences. From each hospital, 20 patients meeting the specifications were selected and assigned at random to subgroups of 5 patients each. The subgroups were then assigned at random to one of the treatment conditions allocated to the hospital to which the subgroup belonged.

Each patient was rated by a panel of judges before and after the treatment combinations were administered. The difference between these two ratings was taken as the criterion of progress. Since some of the treatments were administered to the subgroups as a unit, the subgroup of five patients was considered to be the experimental unit, rather than the individual patient. The mean criterion score for each subgroup is given in Table 8.7-2 to the right of the corresponding symbol for the treatment combination. Each of these means is considered to be a single observation for purposes of the analysis of variance.

Summary tables and details of the computations are given in Table 8.7-3. Cell totals for the treatment combinations appear in part i; these totals are obtained by combining the data from the two replications. Each of these cell totals is considered to be the sum of two observations. The numerical values for individual comparisons corresponding to the main effects and the interactions are computed in part ii. Here the entries in the upper half

of column 1 are obtained from the total column by summing successive pairs of values in the total column; the entries in the lower half of column 1 are obtained by taking the difference between successive pairs of values in the total column. Column 2 is obtained from column 1 by an analogous procedure, and similarly column 3 is obtained from column 2. The entries

Table 8.7-3 Summary of Numerical Analysis

	b_0		b_1		Total
	c_0	c_1	c_0	c_1	
(i) a_0	5	15	25	40	85
a_1	20	40	10	20	90
	25	55	35	60	175

	Treatment combinations				
	Total	(1)	(2)	(3)	
(ii) (1)	5	25	60	175	G
a	20	35	115	5	$A = A_1 - A_0$
b	25	55	0	15	$B = B_1 - B_0$
ab	10	60	5	-75	AB
c	15	15	10	55	$C = C_1 - C_0$
ac	40	-15	5	5	AC
bc	40	25	-30	-5	BC
abc	20	-20	-45	-15	ABC

(iii) $SS_A = (5)^2/16 = 1.57$	$SS_{AC} = (5)^2/16 = 1.57$
$SS_B = (15)^2/16 = 14.06$	$SS_{BC} = (-5)^2/16 = 1.57$
$SS_C = (55)^2/16 = 189.06$	$SS_{ABC} = (-15)^2/16 = 14.06$
$SS_{AB} = (-75)^2/16 = 351.57$	
$SS_{\text{hospitals}} = (30^2 + 39^2 + 50^2 + 56^2)/4 - (175)^2/16 = 100.19$	
$SS_{\text{total}} = 2605 - (175)^2/16 = 690.94$	

in column 3 are the values of the comparisons. The corresponding sums of squares are given in part iii. The sum of squares for the main effect of factor A has the general definition

$$SS_A = \frac{C_A^2}{2^k r},$$

where C_A = comparison corresponding to main effect of factor A ,
 k = number of factors,
 r = number of replications.

In this case C_A is the entry in column 3 in row a , which is 5. For this case $r = 2$, and $k = 3$; hence

$$SS_A = \frac{(5)^2}{2^3(2)} = \frac{25}{16} = 1.57.$$

The analysis of variance for these data is given in Table 8.7-4. The within-hospital sum of squares is obtained by subtracting the between-hospital sum of squares from the total sum of squares. The within-hospital residual is obtained from the relation

$$\begin{aligned} SS_{\text{res}} &= SS_{\text{w. hospital}} - (\text{sum of main effects and two-factor interactions}) \\ &= 590.75 - 559.40 = 31.35. \end{aligned}$$

Although the ABC interaction is completely confounded with between-hospital differences within a single replication, by including information from the two replications the ABC component may be estimated.

Table 8.7-4 Analysis of Variance

Source	SS	df	MS	F
Between hospitals	100.19	3		
Replications	85.56	1		
Residual (b)	14.63	2		
Within hospitals	590.75	12		
A Drug A	1.57	1	1.57	—
B Drug B	14.06	1	14.06	2.69
C Psychotherapy	189.06	1	189.06	36.22**
AB	351.57	1	351.57	67.35**
AC	1.57	1	1.57	—
BC	1.57	1	1.57	—
Residual (w)	31.35	6	5.22	

$$**F_{.99}(1,6) = 13.74$$

The within-hospital residual mean square is used as the denominator for all tests on within-hospital effects. The F tests show no significant interactions involving factor C (psychotherapy), but a significant main effect. It may be concluded from this information that the effect of psychotherapy is independent of the effect of the drugs; further, the groups given psychotherapy showed significantly greater improvement than did groups which were not given psychotherapy.

The interaction between the drugs is seen to be highly significant. The following summary data obtained from part i of Table 8.7-3 are useful in the interpretation of this interaction:

	b_0	b_1
a_0	20	65
a_1	60	30

These data indicate that the use of both drugs simultaneously is not better than the use of either drug alone. The test on the comparison between the

effects (with the exception of ABC) with the hospitals receiving the same set of treatment combinations. Direct computation of the residual sum of squares is illustrated in Table 8.7-5. Since hospitals 1 and 3 had the same set of treatment conditions, the data at the left in part i provide three of the six degrees of freedom of the residual term. If the four treatment combinations common to hospitals 1 and 3 are considered as four levels of a single factor, then the interaction of the levels of this factor with the hospital factor (defined by hospitals 1 and 3) is part of the sum of squares for residuals. This interaction is computed at the left in part ii. An analogous interaction term is obtained from the data at the right in part i.

8.8 Numerical Example of 2 × 2 × 2 Factorial Experiment in Blocks of Size 4 (Repeated Measures)

The purpose of this experiment was to evaluate the preferences for advertisements made up by varying the size (factor A), the style of type (factor B), and the color (factor C). The definitions of the levels of these three factors are given in Table 8.8-1.

Table 8.8-1 Definition of Factors

Factor	Level	Definition
Size (A)	a_0	Small
	a_1	Large
Style (B)	b_0	Gothic
	b_1	Roman
Color (C)	c_0	Green
	c_1	Blue

The experimenter desired to have within-subject estimates on all main effects and interactions; however, the task of having each subject judge all eight combinations was not considered to be experimentally feasible. Loss of interest on the part of the subjects and excessive time demands (as indicated by a pilot study) ruled against the procedure of having each subject judge all combinations. The experimenter was willing to sacrifice precision with respect to the three-factor interaction in order to keep the number of judgments an individual had to make down to four. The plan outlined in Table 8.8-2 was chosen for use.

In order to keep this illustrative example simple, it will be assumed that a sample of six subjects was used in the experiment. (In practice, this sample size would be too small.) The subjects were divided at random into two groups. Individuals within each group judged only four of the eight different make-ups. The combinations of factors judged by group I satisfied the relation $x_1 + x_2 + x_3 = 1 \pmod{2}$, and the combinations judged by group II satisfied the relation $x_1 + x_2 + x_3 = 0 \pmod{2}$. In terms of the specialized notation system for 2^k factorial experiments, group

I judged treatment combinations represented by symbols having an odd number of letters; group II judged the remaining treatment combinations. In this design the *ABC* comparison is completely confounded with differences between groups; within-subject estimates are available on all other factorial effects.

Table 8.8-2 Outline of Plan and Basic Data

Person

Group I

Group II

(i)

Person	a	b	c	abc	Total	Person	(1) ab	ac	bc	Total
1	16	8	2	8	34	4	10	12	8	3
2	10	4	3	7	24	5	11	16	10	5
3	9	3	0	5	17	6	4	7	7	2
Total	35	15	5	20	75	Total	25	35	25	10

(ii)

	Treat comb.: a b c abc (1) ab ac bc								Comparison	
	Cell totals:	35	15	5	20	25	35	25		10
G		+	+	+	+	+	+	+	+	170
A		+	-	-	+	-	+	+	-	60
B		-	+	-	+	-	+	-	+	-10
C		-	-	+	+	-	-	+	+	-50
AB		-	-	+	+	+	+	-	-	0
AC		-	+	-	+	+	-	+	-	0
BC		+	-	-	+	+	-	-	+	10
ABC		+	+	+	+	-	-	-	-	-20

(iii)

SS _A = (60) ² /24 = 150.00	SS _{AB} = 0	SS _{BC} = (10) ² /24
SS _B = (-10) ² /24 = 4.17	SS _{AC} = 0	= 4.17
SS _C = (-50) ² /24 = 104.17	SS _{ABC} = (-20) ² /24 = 16.67	
SS _{subj} = (34 ² + 24 ² + 17 ² + 33 ² + 42 ² + 20 ²)/4 - (170) ² /24 = 114.33		
SS _{total} = (16 ² + 8 ² + ... + 7 ² + 2 ²) - (170) ² /24 = 409.83		

The order in which a subject judged a particular combination was randomized independently for each subject. Order could, however, have been controlled by means of a Latin square. (Plan 10.7-7 utilizes the Latin-square principle with what is essentially the design given in Table 8.8-2. The number of subjects per group would have to be a multiple of 4 in order to use the Latin square.)

The comparisons corresponding to the main effects and interactions of the factors are obtained at the right in part ii of Table 8.8-2. The method by which the *ABC* comparison is obtained shows clearly that this comparison is completely confounded with differences between the two groups. The sums of squares corresponding to the factorial effects, between-subject variation, and total variation are computed in part iii. The over-all analysis of variance is summarized in Table 8.8-3.

The sum of squares due to the within-subject variation is obtained from the relation

$$\begin{aligned} SS_{w, \text{subj}} &= SS_{\text{total}} - SS_{\text{subj}} \\ &= 409.83 - 114.33 = 295.50. \end{aligned}$$

The residual sum of squares is given by

$$\begin{aligned} SS_{\text{res}} &= SS_{w, \text{subj}} - (\text{sum of main effects and 2-factor interactions}) \\ &= 295.50 - 262.51 = 32.99. \end{aligned}$$

Assuming that factors *A*, *B*, and *C* are fixed, the residual term is the proper denominator for all within-subject effects. The proper denominator for

Table 8.8-3 Summary of Analysis

Source of variation	SS	df	MS	<i>F</i>
Between subjects	114.33	5		
<i>ABC</i> (groups)	16.67	1	16.67	
Subjects within groups	97.67	4	24.42	
Within subjects	295.50	18		
<i>A</i> Size	150.00	1	150.00	54.54**
<i>B</i> Type style	4.17	1	4.17	1.51
<i>C</i> Color	104.17	1	104.17	37.88**
<i>AB</i>	0	1	0	
<i>AC</i>	0	1	0	
<i>BC</i>	4.17	1	4.17	1.51
Residual	32.99	12	2.75	

$$** F_{.99}(1,12) = 9.33$$

between-subject effects is the mean square for subjects within groups. In this design the *ABC* interaction is a between-subject effect—but this interaction is completely confounded with differences among groups. Generally there would be little interest in testing the *ABC* interaction. (In this case the *F* ratio would be less than 1.)

For the within-subject data in Table 8.8-3, none of the interactions is statistically significant. This result implies that the main effects (if any) operate independently; i.e., the main effects are additive. The tests made in Table 8.8-3 indicate that the size and color main effects are statistically significant. Inspection of the summary data in Table 8.8-2 indicates that large size (*a*₁) is preferred no matter which of the styles or colors is used. Similarly, green (*c*₀) is preferred to blue no matter which size or style of type is used. There is no statistically significant difference between the two styles of type used in the experiment; there is, however, a slight preference for the Gothic (*b*₀).

If one were to make an over-all recommendation with respect to the most preferred make-up of the advertising copy, the large size in the green color

would be the best combination. The two type styles are, statistically, equally good, but there is a slight observed difference in favor of the Gothic.

Returning to structural considerations underlying the analysis of this design, the residual term in the analysis consists of the following pooled interactions:

Subject \times treatments (within group I)	6
Subject \times treatments (within group II)	6
Residual	12

The residual term in Table 8.8-3 may be computed directly from the interaction terms given above. The four treatment combinations assigned to the

Table 8.8-4 Analysis of General Case of Design in Table 8.8-2

Source of variation	df
Between subjects	$2nr - 1$
Replications	$r - 1$
Groups within replications	r
<i>ABC</i>	1
Residual (groups)	$r - 1$
Subjects within groups	$2r(n - 1)$
Within subjects	$6nr$
<i>A</i>	1
<i>B</i>	1
<i>C</i>	1
<i>AB</i>	1
<i>AC</i>	1
<i>BC</i>	1
Residual (within subject)	$6(nr - 1)$

subjects within group I are considered to be four levels of a single factor; the subject by treatment interaction is computed by means of the usual computational formulas for a two-factor interaction.

The analysis of the general case of a $2 \times 2 \times 2$ factorial experiment in blocks of size 4 with repeated measures is outlined in Table 8.8-4. In this analysis there are n subjects in each group, and the experiment is replicated r times. The analysis of the design in Table 8.8-3 is a special case of the more general design in Table 8.8-4. In the special case $n = 3$, and $r = 1$. In a design having more than one replication, the *ABC* interaction may be tested by using the subjects within groups in the denominator of an F ratio. This test will generally be considerably less sensitive than tests on the main effects and two-factor interactions.

8.9 Designs for 3×3 Factorial Experiments

The nine treatment combinations in this experiment may be partitioned into three sets in such a way that differences between sets form one of the components of the $A \times B$ interaction. The design outlined in Table 8.9-1

gives two such partitions. In replication 1, the AB (or J) component is used to define the sets (blocks); in replication 2, the AB^2 (or I) component is used to define the sets (blocks).

The treatment combinations in block 1 satisfy the relation $x_1 + x_2 = 0 \pmod{3}$; the treatment combinations in blocks 2 and 3 satisfy the respective relations $x_1 + x_2 = 1 \pmod{3}$ and $x_1 + x_2 = 2 \pmod{3}$. Hence differences between blocks within replication 1 are completely confounded with the AB (or J) component of $A \times B$. That is, a difference between two block totals in replication 1 is simultaneously an estimate of block differences as well as differences which form part of the AB component. (A formal proof of this is given later in this section.)

In replication 2, each of the blocks contains treatment combinations which satisfy the relation $x_1 + 2x_2 = i \pmod{3}$. For example, the treatment

Table 8.9-1 3×3 Factorial Experiment in Blocks of Size 3

Replication 1			Replication 2		
Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
(00)	(01)	(02)	(00)	(02)	(01)
(12)	(10)	(20)	(11)	(10)	(12)
(21)	(22)	(11)	(22)	(21)	(20)
Component confounded AB (or J)			AB^2 (or I)		

combinations in block 5 all satisfy the relation $x_1 + 2x_2 = 2 \pmod{3}$. Hence differences between blocks within replication 2 are completely confounded with the AB^2 (or I) component of $A \times B$.

Assuming one observation per cell, the analysis takes the form given in Table 8.9-2. Within-block information on the main effects is available from both replications. Within-block information on the AB (or J) component is available only from replication 2; within-block information on the AB^2 (or I) component is available only from replication 1. Since within-block information on the components of $A \times B$ is available from only one of the two replications, the relative within-block information on $A \times B$ is said to be $\frac{1}{2}$. For main effects, however, the relative within-block information is $\frac{2}{2} = 1$, since both replications provide within-block information on the main effects. The design in Table 8.9-1 is balanced in the sense that the same amount of relative within-block information is provided on each of the components of $A \times B$.

The denominator of F ratios for within-block effects is the pooled interaction of the main effects with the replications. The two components of the $A \times B$ interaction are generally combined, and the combined interaction (with four degrees of freedom) is tested as a unit. If, however, the individual

components are meaningful in terms of the experimental variables, separate tests on the components may be made. In most cases the components of $A \times B$ that are confounded with block effects are not tested. There are, however, techniques for combining the between-block information on the interaction with the within-block information to obtain an over-all estimate of the interaction. Such techniques are said to recover the between-block information on the interaction effects; these techniques will not be considered in this section.

The block in this design may be an individual subject—in this case each subject is observed under all treatment combinations assigned to a given block. The block may be a group of three subjects—in this case each

Table 8.9-2 Analysis of Design in Table 8.9-1

Source		df	E(MS)
Between blocks		5	
Replications		1	
AB (from rep 1)		2	
AB^2 (from rep 2)		2	
Within blocks		12	
A		2	$\sigma_e^2 + 6\sigma_\alpha^2$
B		2	$\sigma_e^2 + 6\sigma_\beta^2$
$A \times B$		4	
AB (from rep 2)	2		$\sigma_e^2 + (\frac{1}{2})(2)\sigma_{\alpha\beta J}^2$
AB^2 (from rep 1)	2		$\sigma_e^2 + (\frac{1}{2})(2)\sigma_{\alpha\beta I}^2$
Residual		4	σ_e^2
$A \times \text{rep}$	2		
$B \times \text{rep}$	2		

subject is observed under the three treatment combinations assigned to the block. The block may be a group of n subjects—in this case the n subjects may be observed under all treatment combinations within a given block. The analysis of this latter design takes the form given in Table 8.9-3.

The expected values of the mean squares have the same general form as those given in Table 8.9-2. The residual again provides an estimate of the within-block experimental error. The $12n - 8$ degrees of freedom for the residual is the pooled interaction of the main effects with the replications and the interaction of the main effects with the subjects within groups. The breakdown of these degrees of freedom is as follows:

Residual	$12n - 8$
$A \times \text{rep}$	2
$B \times \text{rep}$	2
$A \times \text{subject within group}$	$12n - 6$
$B \times \text{subject within group}$	$12n - 6$

A formal algebraic proof will now be outlined to demonstrate that the AB component in the design in Table 8.9-1 is confounded with block effects within replication 1 but is free of such confounding within replication 2. This proof assumes that an observation under treatment combination ab_{ij} in block k provides an estimate (disregarding the experimental error) of the sum of the parameters on the right-hand side of the following expression,

$$X_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \pi_k,$$

where the symbol π_k designates the effect of block k . Assuming that factors A and B are fixed, $\sum \alpha_i = 0$, $\sum \beta_j = 0$, $\sum_i (\alpha\beta)_{ij} = 0$, and $\sum_j (\alpha\beta)_{ij} = 0$. These restrictions on the above model follow directly from the basic definition of treatment effects. Interactions with block effects do not appear on the

Table 8.9-3 Analysis of 3×3 Factorial Experiment in Blocks of Size 6 (Repeated Measurements)

Source		df
Between subjects		$6n - 1$
Groups		5
Replications	1	
AB (from rep 1)	2	
AB^2 (from rep 2)	2	
Subjects within groups		$6(n - 1)$
Within subjects		$12n$
A		2
B		2
AB (from rep 2)		2
AB^2 (from rep 1)		2
Residual		$12n - 8$

right-hand side of the model, hence the implicit assumption that such effects either do not exist or that such effects are negligible relative to the magnitude of the other effects.

The set of treatment combinations which satisfy the relation $x_1 + x_2 = i$ define the set J_i . It will be convenient to define the sum of the interaction effects in the set J_0 by the symbol $3(\alpha\beta)_{J_0}$. Thus

$$3(\alpha\beta)_{J_0} = (\alpha\beta)_{00} + (\alpha\beta)_{12} + (\alpha\beta)_{21}.$$

Similarly, by definition,

$$3(\alpha\beta)_{J_1} = (\alpha\beta)_{01} + (\alpha\beta)_{10} + (\alpha\beta)_{22},$$

$$3(\alpha\beta)_{J_2} = (\alpha\beta)_{02} + (\alpha\beta)_{11} + (\alpha\beta)_{20}.$$

In terms of the basic model and the specialized definitions, for observations on treatment combinations in the set J_0 in replication 1 of the design in

Table 8.9-1,

$$X_{00} \doteq \mu + \alpha_0 + \beta_0 + (\alpha\beta)_{00} + \pi_1$$

$$X_{12} \doteq \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} + \pi_1$$

$$X_{21} \doteq \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} + \pi_1$$

$$J_0. \doteq 3\mu + 3(\alpha\beta)_{J_0} + 3\pi_1$$

The symbol $J_0.$ is used to designate the sum of all observations on treatment combinations in the set $J_0.$ Similarly,

$$J_{1.} \doteq 3\mu + 3(\alpha\beta)_{J_1} + 3\pi_2,$$

$$J_{2.} \doteq 3\mu + 3(\alpha\beta)_{J_2} + 3\pi_3.$$

The sum of squares for the AB component of $A \times B$ is obtained from differences between these J totals. For example,

$$J_{0.} - J_{1.} \doteq 3(\alpha\beta)_{J_0} - 3(\alpha\beta)_{J_1} + 3\pi_1 - 3\pi_2.$$

Each of the other possible comparisons among the J 's also includes block effects. Hence, from the data in replication 1, SS_{AB} is completely confounded with differences among blocks.

In contrast, for replication 2,

$$X_{00} \doteq \mu + \alpha_0 + \beta_0 + (\alpha\beta)_{00} + \pi_4$$

$$X_{12} \doteq \mu + \alpha_1 + \beta_2 + (\alpha\beta)_{12} + \pi_6$$

$$X_{21} \doteq \mu + \alpha_2 + \beta_1 + (\alpha\beta)_{21} + \pi_5$$

$$J_{0.} \doteq 3\mu + 3(\alpha\beta)_{J_0} + \pi_4 + \pi_5 + \pi_6$$

Similarly, for replication 2,

$$J_{1.} \doteq 3\mu + 3(\alpha\beta)_{J_1} + \pi_4 + \pi_5 + \pi_6,$$

$$J_{2.} \doteq 3\mu + 3(\alpha\beta)_{J_2} + \pi_4 + \pi_5 + \pi_6.$$

Differences between J totals in replication 2 are free of block effects. For example,

$$J_{0.} - J_{1.} \doteq 3(\alpha\beta)_{J_0} - 3(\alpha\beta)_{J_1}.$$

Hence within-block information on the AB components of $A \times B$ may be obtained from replication 2.

To show that replication 1 provides within-block information on the AB^2 (or I) component of $A \times B$,

$$X_{00} \doteq \mu + \alpha_0 + \beta_0 + (\alpha\beta)_{00} + \pi_1$$

$$X_{11} \doteq \mu + \alpha_1 + \beta_1 + (\alpha\beta)_{11} + \pi_3$$

$$X_{22} \doteq \mu + \alpha_2 + \beta_2 + (\alpha\beta)_{22} + \pi_2$$

$$I_{0.} \doteq 3\mu + 3(\alpha\beta)_{I_0} + \pi_1 + \pi_2 + \pi_3$$

Similarly, for replication 1,

$$I_{1.} \doteq 3\mu + 3(\alpha\beta)_{I_1} + \pi_1 + \pi_2 + \pi_3,$$

$$I_{2.} \doteq 3\mu + 3(\alpha\beta)_{I_2} + \pi_1 + \pi_2 + \pi_3.$$

Since each I total contains the same block effects, differences between these totals will be free of block effects. Hence information on the I component of $A \times B$ obtained from replication 1 will not be confounded with block effects.

For the general case of a design having r replications (one observation per cell) of the design in Table 8.9-1, with $r/2$ replications of the form of replication 1 and $r/2$ of the form of replication 2, the sum of squares for the J component of $A \times B$ is

$$SS'_{AB(J)} = \frac{\sum J_i^2}{3(r/2)} - \frac{(\sum J_i)^2}{9(r/2)},$$

where J totals are restricted to the replications in which the J component of $A \times B$ is free of block effects. Similarly, the sum of squares for the I components of $A \times B$ is given by

$$SS'_{AB(I)} = \frac{\sum I_i^2}{3(r/2)} - \frac{(\sum I_i)^2}{9(r/2)},$$

where the I totals are restricted to replications in which the I component of $A \times B$ is free of block effects.

If the symbol AB_{ij} represents the sum of all observations on treatment combination ab_{ij} , then this sum will not be free of block effects, since all treatment combinations do not appear in each of the blocks. An adjusted sum which is free of block effects is given by

$$AB'_{ij} = \frac{A_i + B_j}{3} + \frac{J_{(i+j)} + I_{(i+2j)}}{\frac{3}{2}} - \frac{G}{3},$$

where A_i = sum of all observations at level a_i ,

B_j = sum of all observations at level b_j ,

$J_{(i+j)}$ = sum of all observations on treatment combinations which satisfy the relation $x_1 + x_2 = i + j \pmod{3}$ in replications in which differences between J 's are free of block effects,

$I_{(i+2j)}$ = sum of all observations on treatment combinations which satisfy the relation $x_1 + 2x_2 = i + 2j \pmod{3}$ in replications in which differences between I 's are free of block effects.

For example, the adjusted total for all observations made under treatment combination ab_{02} is

$$AB'_{02} = \frac{A_0 + B_2}{3} + \frac{J_2 + I_1}{\frac{3}{2}} - \frac{G}{3}.$$

In terms of the general linear model, this last expression estimates the parameters on the right-hand side of the following expression:

$$AB'_{02} = r\mu + r\alpha_0 + r\beta_2 + r(\alpha\beta)_{J_2} + r(\alpha\beta)_{I_1} \\ + \frac{\text{sum of all block effects}}{3}.$$

Since all adjusted totals will contain this last term, differences between the adjusted totals will be free of block effects. To demonstrate that AB'_{02} actually does estimate the parameters on the right-hand side,

$$\frac{A_0}{3} \doteq r\mu + r\alpha_0 + \frac{1}{3}(\text{sum of all block effects})$$

$$\frac{B_2}{3} \doteq r\mu + r\beta_2 + \frac{1}{3}(\text{sum of all block effects})$$

$$\frac{J_2}{\frac{3}{2}} \doteq r\mu + r(\alpha\beta)_{J_2} + \frac{2}{3}(\text{sum of blocks in which } J \text{ is free of block effects})$$

$$\frac{I_1}{\frac{3}{2}} \doteq r\mu + r(\alpha\beta)_{I_1} + \frac{2}{3}(\text{sum of blocks in which } I \text{ is free of block effects})$$

$$\frac{-G}{3} \doteq -3r\mu - (\text{sum of all block effects})$$

$$AB'_{02} \doteq r\mu + r\alpha_0 + r\beta_2 + r(\alpha\beta)_{J_2} + r(\alpha\beta)_{I_1} + \frac{\text{sum of all block effects}}{3}$$

In making comparisons among the means of treatment combinations, one uses the adjusted totals to obtain the means. The effective number of observations in each of these adjusted means is less than r , since information on the I and J components of the $A \times B$ interaction is obtained from only one-half of the replications. For most practical purposes, however, the adjusted cell means may be handled as if they were based upon r replications. The adjusted mean for cell ab_{ij} is

$$\overline{AB'_{ij}} = \frac{AB'_{ij}}{r}.$$

If this design had r replications with n observations per cell, AB'_{ij} would have the definition given above but

$$\overline{AB'_{ij}} = \frac{AB'_{ij}}{nr}.$$

8.10 Numerical Example of 3×3 Factorial Experiment in Blocks of Size 3

In the experiment to be described it was desired to have within-subject information on all factorial effects in a 3×3 factorial experiment. However, it was not experimentally feasible to have each subject observed under each of the nine treatment combinations. Each person could be observed under three treatment combinations. The design outlined in Table 8.9-1 was selected for use.

The experimenter wished to have two observations under each treatment combination. A random sample of 12 subjects was obtained from a specified population; the sample was divided at random into six groups of 2

subjects each. The groups were then assigned at random to the blocks in Table 8.10-1. The 2 subjects within each group were observed under each of the three treatment combinations in the assigned block. The order in which a subject was observed under a treatment condition was randomized for each subject. The data obtained from this experiment are summarized in Table 8.10-1.

Table 8.10-1 Summary of Observed Data

Block	Person				Total
1	1	(00)	(12)	(21)	$36 = P_1$ $14 = P_2$ $50 = G_1$
	2	14	7	15	
		6	3	5	
		(01)	(10)	(22)	
2	3	3	4	10	$17 = P_3$ $43 = P_4$ $60 = G_2$
	4	7	6	30	
		(02)	(11)	(20)	
	5	5	15	7	
3	6	5	5	3	$27 = P_5$ $13 = P_6$ $40 = G_3$
		(00)	(11)	(22)	
	7	10	10	15	
4	8	15	20	25	$35 = P_7$ $60 = P_8$ $95 = G_4$
		(02)	(10)	(21)	
	9	3	5	12	$20 = P_9$ $35 = P_{10}$ $55 = G_5$
	10	7	10	18	
		(01)	(12)	(20)	
	11	6	7	5	
6	12	4	3	5	$18 = P_{11}$ $12 = P_{12}$ $30 = G_6$
					$180 = R_2$

In the row having person 1 at the left, the entries 14, 7, and 15 are the respective observations (with order randomized) on person 1 under treatment combinations (00), (12), and (21). The sum of these three observations on person 1 is denoted by the symbol P_1 . The symbols G_i denote the totals for the six observations made within a group (block). The symbols R_j denote the totals of the 18 observations made within a single replication. Additional data required for the computation of the sums of squares in the analysis are given in Table 8.10-2.

The sums of squares for all effects except the $A \times B$ interaction are computed in the same manner as that used in any factorial experiment having repeated measures. Data required for computation of all effects except

the components of $A \times B$ are summarized in parts ii and iii of Table 8.10-2. From the data in part iii, the between-subject sum of squares is given by

$$\frac{\Sigma P^2}{3} - \frac{G^2}{36} = 3815.33 - 3025.00 = 790.33.$$

The within-subject sum of squares is given by

$$\Sigma \Sigma X^2 - \frac{\Sigma P^2}{3} = 4494 - 3815.33 = 678.67.$$

Data required for the sums of squares for groups, replications, and the main effects of factors A and B are also given in part iii.

Table 8.10-2 Summary Data Required for Analysis

		Replication 1				Replication 2			
		b_0	b_1	b_2		b_0	b_1	b_2	
(i)	a_0	20	10	10		25	10	10	
	a_1	10	20	10		15	30	10	
	a_2	40	10	20	$-80 = I_0$	50	10	30	-95
	a_0	50	20	10	$-40 = I_1$	65	25	10	-55
	a_1	60	10	20	$-30 = I_2$	65	15	30	-30
		150		150		180		180	
		Combined replications							
		b_0	b_1	b_2	Total				
(ii)	a_0	45	20	20	$85 = A_0$				
	a_1	25	50	20	$95 = A_1$				
	a_2	20	50	80	$150 = A_2$				
		90	120	120	$330 = G$				
		B_0	B_1	B_2					
(iii)		$G^2/9nr = 3025.00$				$(\Sigma P^2)/3 = 3815.33$			
		$\Sigma \Sigma X^2 = 4494$				$(\Sigma G_i^2)/3n = 3441.67$			
		$(\Sigma A^2)/3nr = 3229.17$				$(\Sigma R^2)/9n = 3050.00$			
		$(\Sigma B^2)/3nr = 3075.00$							

From the data in part i, both the between- and within-subject components of the $A \times B$ interaction may be computed. From the totals at the right of replication 1, the within-subject information on the I component of $A \times B$ is given by

$$\begin{aligned} SS_{AB(I)} &= SS_{AB^2} = \frac{\Sigma I_i^2}{3n} - \frac{(\Sigma I_i)^2}{9n} \\ &= \frac{80^2 + 40^2 + 30^2}{3(2)} - \frac{(150)^2}{18} = 233.33. \end{aligned}$$

From the totals at the left of replication 2, the within-subject information on the J component of $A \times B$ is given by

$$\begin{aligned} SS_{AB(J)} = SS_{AB} &= \frac{\sum J_i^2}{3n} - \frac{(\sum J_i)^2}{9n} \\ &= \frac{50^2 + 65^2 + 65^2}{6} - \frac{(180)^2}{18} = 25.00. \end{aligned}$$

Thus within-subject information on the components of the $A \times B$ interaction is obtained separately from each of the replications—one component from replication 1, the other from replication 2.

Table 8.10-3 Summary of Analysis

Source	SS	df	MS	F
Between subjects	790.33	11		
Groups (blocks)	416.67	5		
Replications	25.00	1		
AB (from rep 1)	33.33	2		
AB^2 (from rep 2)	358.33	2		
Subjects within groups	373.66	6	62.28	
Within subjects	678.67	24		
A	204.17	2	102.08	9.82**
B	50.00	2	25.00	2.41
$A \times B$ (adjusted)	258.33	4	64.58	6.22**
AB (from rep 2)	25.00	2		
AB^2 (from rep 1)	233.33	2		
Residual	166.17	16	10.39	

** $F_{.99}(2,16) = 6.23$; $F_{.99}(4,16) = 4.77$

The between-subject information on the J component of $A \times B$ is obtained from the totals at the left of replication 1.

$$SS_{AB(J)} = SS_{AB} = \frac{40^2 + 50^2 + 60^2}{6} - \frac{(150)^2}{18} = 33.33.$$

The between-subject information on the I component of $A \times B$ is obtained from the totals to the right of replication 2.

$$SS_{AB(I)} = SS_{AB^2} = \frac{95^2 + 55^2 + 30^2}{6} - \frac{(180)^2}{18} = 358.33.$$

The residual sum of squares in Table 8.10-3 is obtained from the relation

$$SS_{\text{res}} = SS_{\text{w. subj}} - SS_A - SS_B - SS_{AB(\text{within})} - SS_{AB^2(\text{within})}.$$

The residual sum of squares may also be computed directly from the interactions of factor A and B with the replications and subjects within groups.

Unless there is an a priori reason for handling the components of the $A \times B$ interaction separately, a single test is made on the combined within-subject components of this interaction. (In cases in which the large cell values in a two-way summary table fall along the diagonal running from upper left to lower right, the AB^2 component will be large relative to the AB component.) In this numerical example the $A \times B$ interaction is statistically significant. The simple effects for both factor A and factor B are between-subject effects; however, approximations to tests on simple effects may be obtained by working with the adjusted cell totals. Differences among the latter totals may be considered to be within-subject effects.

To illustrate the computation of the adjusted cell totals, for treatment combination ab_{02}

$$\begin{aligned} AB'_{02} &= \frac{A_0 + B_2}{3} + \frac{I_1 + J_2}{\frac{3}{2}} - \frac{G}{3} \\ &= \frac{85 + 120}{3} + \frac{40 + 50}{\frac{3}{2}} - \frac{330}{3} = 18.33. \end{aligned}$$

(Note that in part i of Table 8.10-2 cell ab_{02} contributes to the I_1 total in replication 1 and to the J_2 total in replication 2.) The unadjusted cell total is 20. As another example, the adjusted total for cell ab_{11} has the form

$$AB'_{11} = \frac{A_1 + B_1}{3} + \frac{I_0 + J_2}{\frac{3}{2}} - \frac{G}{3}.$$

Comparisons among the adjusted cell totals use the within-subject residual mean square as an error term. For approximate tests, the number of effective observations in each cell is considered to be $nr = 4$.

To illustrate the procedures for tests on adjusted cell totals,

$$F = \frac{(AB'_{02} - AB'_{22})^2}{2nr \text{ MS}_{\text{res}}}.$$

The adjusted total AB'_{02} was found to be 18.33. The adjusted total AB'_{22} is given by

$$AB'_{22} = \frac{A_2 + B_2}{3} + \frac{I_0 + J_1}{\frac{3}{2}} - \frac{G}{3}.$$

(Note that ab_{22} contributes to the I_0 total in replication 1 and to the J_1 total in replication 2.)

$$\begin{aligned} AB'_{22} &= \frac{150 + 120}{3} + \frac{80 + 65}{\frac{3}{2}} - \frac{330}{3} \\ &= 76.67. \end{aligned}$$

Thus,

$$F = \frac{(18.33 - 76.67)^2}{2(2)(2)(10.39)} = 40.95.$$

The critical value for a .05-level (a priori) test is $F_{.95}(1,16) = 4.49$.

8.11 Designs for $3 \times 3 \times 3$ Factorial Experiments

The 27 treatment combinations in a $3 \times 3 \times 3$ factorial experiment may be divided into sets in several different ways and balance still be maintained with respect to main effects and two-factor interactions. Each of the following components of the $A \times B \times C$ interactions can serve to define such balanced sets:

$A \times B \times C$	df
ABC	8
ABC^2	2
AB^2C	2
AB^2C^2	2

For example, the sets formed by using the ABC component are as follows

$x_1 + x_2 + x_3 = 0 \pmod{3}$	$x_1 + x_2 + x_3 = 1 \pmod{3}$	$x_1 + x_2 + x_3 = 2 \pmod{3}$
(000)	(001)	(002)
(012)	(010)	(011)
(021)	(022)	(020)
(102)	(100)	(101)
(111)	(112)	(110)
(120)	(121)	(122)
(201)	(202)	(200)
(210)	(211)	(212)
(222)	(220)	(221)

A design consisting of these three blocks would provide within-block information on all effects except the ABC component of the three-factor interaction. With only a single replication (assuming only one observation per cell) no estimate of experimental error is available.

If the experiment were to be replicated by using the same block structure as that given above, the design would lack balance with respect to the components of the three-factor interaction. That is, within-block information would be available on three of the four components of this interaction, but no within-block information would be available on the ABC component. A design which has balance with respect to the components of the three-factor interaction may be constructed by means of the scheme outlined in Table 8.11-1. A minimum of four replications is required for balance. The blocks within each of the replications are constructed by means of relations associated with a component of the three-factor interaction; each of the replications involves a different component. Hence within-block information on all of the three-factor components is available from some three of the four replications.

The structure of the blocks defined by the design in Table 8.11-1 is given in Table 8.11-2. Blocks in replication 1 are balanced with respect to all main effects and all two-factor interactions. These blocks are also balanced

Table 8.11-1 Construction of Blocks of Size 9

Replication	Block	Component confounded	Defining relation (mod 3)
1	1	$(ABC)_0$	$x_1 + x_2 + x_3 = 0$
	2	$(ABC)_1$	$x_1 + x_2 + x_3 = 1$
	3	$(ABC)_2$	$x_1 + x_2 + x_3 = 2$
2	4	$(ABC^2)_0$	$x_1 + x_2 + 2x_3 = 0$
	5	$(ABC^2)_1$	$x_1 + x_2 + 2x_3 = 1$
	6	$(ABC^2)_2$	$x_1 + x_2 + 2x_3 = 2$
3	7	$(AB^2C)_0$	$x_1 + 2x_2 + x_3 = 0$
	8	$(AB^2C)_1$	$x_1 + 2x_2 + x_3 = 1$
	9	$(AB^2C)_2$	$x_1 + 2x_2 + x_3 = 2$
4	10	$(AB^2C^2)_0$	$x_1 + 2x_2 + 2x_3 = 0$
	11	$(AB^2C^2)_1$	$x_1 + 2x_2 + 2x_3 = 1$
	12	$(AB^2C^2)_2$	$x_1 + 2x_2 + 2x_3 = 2$

Table 8.11-2 Design Corresponding to Table 8.11-1

Replication 1			Replication 2		
Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
(000)	(001)	(002)	(000)	(002)	(001)
(012)	(010)	(011)	(011)	(010)	(012)
(021)	(022)	(020)	(022)	(021)	(020)
(102)	(100)	(101)	(101)	(100)	(102)
(111)	(112)	(110)	(112)	(111)	(110)
(120)	(121)	(122)	(120)	(122)	(121)
(201)	(202)	(200)	(202)	(201)	(200)
(210)	(211)	(212)	(210)	(212)	(211)
(222)	(220)	(221)	(221)	(220)	(222)
ABC			ABC^2		
Replication 3			Replication 4		
Block 7	Block 8	Block 9	Block 10	Block 11	Block 12
(000)	(001)	(002)	(000)	(002)	(001)
(011)	(012)	(010)	(012)	(011)	(010)
(022)	(020)	(021)	(021)	(020)	(022)
(102)	(100)	(101)	(101)	(100)	(102)
(110)	(111)	(112)	(110)	(112)	(111)
(121)	(122)	(120)	(122)	(121)	(120)
(201)	(202)	(200)	(202)	(201)	(200)
(212)	(210)	(211)	(211)	(210)	(212)
(220)	(221)	(222)	(220)	(222)	(221)
AB^2C			AB^2C^2		

with respect to all components of the three-factor interaction except the ABC component. To illustrate the balance with respect to the ABC^2 component within block 1, the nine treatment combinations in this block may be divided into the following sets:

$(ABC^2)_0$	$(ABC^2)_1$	$(ABC^2)_2$
(000)	(021)	(012)
(120)	(111)	(102)
(210)	(201)	(222)

Differences between the totals for these sets form part of the ABC^2 component.

**Table 8.11-3 Analysis of Design Having Block Size 9
(One Observation per Cell)**

Source of variation	df	E(MS)
Between blocks	11	
Replications	3	
Blocks within reps	8	
ABC (from rep 1)	2	
ABC^2 (from rep 2)	2	
AB^2C (from rep 3)	2	
AB^2C^2 (from rep 4)	2	
Within blocks	96	
A	2	$\sigma_e^2 + 36\sigma_\alpha^2$
B	2	$\sigma_e^2 + 36\sigma_\beta^2$
C	2	$\sigma_e^2 + 36\sigma_\gamma^2$
$A \times B$	4	$\sigma_e^2 + 12\sigma_{\alpha\beta}^2$
$A \times C$	4	$\sigma_e^2 + 12\sigma_{\alpha\gamma}^2$
$B \times C$	4	$\sigma_e^2 + 12\sigma_{\beta\gamma}^2$
$(A \times B \times C)'$ ($\frac{3}{4}$)	8	$\sigma_e^2 + (\frac{3}{4})(4)\sigma_{\alpha\beta\gamma}^2$
ABC (omit rep 1)	2	
ABC^2 (omit rep 2)	2	
AB^2C (omit rep 3)	2	
AB^2C^2 (omit rep 4)	2	
Residual	70	σ_e^2
Main effects \times reps	18	
2-factor int \times reps	36	
3-factor int \times reps	16	

Assuming one observation per cell, the analysis of this design has the form given in Table 8.11-3. The computation of the sums of squares for main effects and two-factor interactions follows the same rules as those of the replicated factorial experiment. In computing the within-block components of $A \times B \times C$, data from replications in which a component is confounded with block effects are not used. For example, to compute the

within-block component of ABC^2 , data from all replications except replication 2 are combined to form a three-way summary table. From this summary table one obtains the totals $(ABC^2)'_0$, $(ABC^2)'_1$, and $(ABC^2)'_2$, where $(ABC^2)'_i$ represents the sum of observations in all cells which satisfy the relation $x_1 + x_2 + 2x_3 = i \pmod{3}$. Then

$$SS'_{ABC^2} = \frac{\Sigma(ABC^2)'_i{}^2}{27} - \frac{[\Sigma(ABC^2)'_i]^2}{81}.$$

As a partial check on the numerical work, $\Sigma(ABC^2)'_i$ is equal to the sum of all observations in replications 1, 3, and 4.

The within-cell residual is most conveniently obtained by subtracting the within-cell treatment effects from the total within-cell variation. The sources of variation which are pooled to form the residual term are shown in Table 8.11-3. These sources may be computed separately. In this design there are only three effective replications on each of the components of the three-factor interaction; thus the $ABC \times$ replications interaction, for example, has four degrees of freedom. Hence the three-factor \times replications interaction has a total of 16 degrees of freedom—4 degrees of freedom for each of the four components of the three-factor interaction.

In making tests on differences between adjusted cell means, adjusted cell totals are required. Such totals are given by

$$\begin{aligned} ABC'_{ijk} = & \frac{1}{9}[A_i + B_j + C_k + (AB)_{i+j} + (AB^2)_{i+2j} + (AC)_{i+k} \\ & + (AC^2)_{i+2k}] + \left(\frac{4}{3}\right)\left(\frac{1}{9}\right)[(ABC)'_{i+j+k} + (ABC^2)'_{i+j+2k} \\ & + (AB^2C)'_{i+2j+k} + (AB^2C^2)'_{i+2j+2k}] - \frac{4}{9}G. \end{aligned}$$

For example, the adjusted cell total for the cell abc_{012} is

$$\begin{aligned} ABC'_{012} = & \frac{1}{9}[A_0 + B_1 + C_2 + (AB)_1 + (AB^2)_2 + (AC)_2 + (AC^2)_1] \\ & + \frac{4}{27}[(ABC)'_0 + (ABC^2)'_2 + (AB^2C)'_1 + (AB^2C^2)'_0] - \frac{4}{9}G. \end{aligned}$$

The symbol $(AB^2C^2)'_0$ designates the sum of treatment combinations which satisfy the relation $x_1 + 2x_2 + 2x_3 = 0$ in replications in which this component is not confounded with blocks. This total is obtained in the course of computing $SS'_{AB^2C^2}$.

Since the two-factor interactions in the design given in Table 8.11-2 are not confounded with block effects, no adjustments are required for cell totals in two-way summary tables. Differences between means computed from such totals will be free of block effects.

Block Size = 3. In dividing the 27 treatment combinations into blocks of size 9, a relation of the general form

$$u = x_1 + u_2x_2 + u_3x_3 = i \pmod{3}$$

was used for each of the replications. The relation u corresponded to some component of the three-factor interaction. In the analysis of the resulting design, the u component was confounded with block effects within the replication having blocks defined by the u relation. The 27 treatment combinations in a $3 \times 3 \times 3$ factorial experiment may be divided into blocks of size 3 by requiring that the treatment combinations within a given block simultaneously satisfy the two relations

$$u = i \pmod{3},$$

$$v = j \pmod{3},$$

where v is some relation other than u .

In the analysis of a replication having blocks formed in this way, the components corresponding to u and v will both be confounded with differences between blocks. In addition the components corresponding to the relations $u + v$ and $u + 2v$ will also be confounded with block effects. The latter components are known as the *generalized interactions*, or *aliases*, of the components corresponding to u and v . To illustrate, suppose that the blocks within a replication are defined by the relations

$$u = x_1 + x_2 + 2x_3 = i \pmod{3},$$

$$v = x_1 + 2x_2 = j \pmod{3}.$$

These relations correspond, respectively, to the ABC^2 and the AB^2 components. One of the generalized interactions, or aliases, is given by

$$u + v = 2x_1 + 3x_2 + 2x_3 = i + j \pmod{3}$$

$$= 2x_1 + 2x_3 = i + j \pmod{3}$$

$$= x_1 + x_3 = 2(i + j) \pmod{3}.$$

(The last line is obtained from the one above by multiplying both sides of the equation by 2 and then reducing coefficients to the modulus 3.) This last relation corresponds to the AC component. The second generalized interaction of u and v is given by

$$u + 2v = 3x_1 + 5x_2 + 2x_3 = i + 2j \pmod{3}$$

$$= 2x_2 + 2x_3 = i + 2j \pmod{3}$$

$$= x_2 + x_3 = 2(i + 2j) \pmod{3}.$$

This last relation corresponds to the BC component.

To summarize this illustrative example, blocks are defined by relations corresponding to the ABC^2 and AB^2 components. The generalized interactions, or aliases, of these components are the AC and BC components.

Hence, in a replication defined by these relations, the analysis would have the following form:

Between blocks	8
AB^2	2
AC	2
BC	2
ABC^2	2
Within blocks	18
A, B, C	2 each
AB, AC^2, BC^2	2 each
ABC, AB^2C, AB^2C^2	2 each

The defining relations in this example may be symbolized by $(ABC^2)_i$ and $(BC^2)_j$; the aliases may be symbolized by $(AC)_{2(i+j)}$ and $(BC)_{2(i+2j)}$. The block defined by $(ABC^2)_0$ and $(BC^2)_1$ is

$$(022), \quad (101), \quad (210).$$

It may be verified that each of the treatment combinations in the blocks also satisfies the relations $(AC)_{2(0+1)} = (AC)_2$ and $(BC)_{2(0+2)} = (BC)_1$.

In general, if two relations are used simultaneously to define blocks, one needs to know the generalized interactions of the components corresponding to these relations in order to carry out the analysis. There is a relatively simple rule for determining the aliases of any two components. Given two components having the general form W and X , their aliases will have the general forms WX and WX^2 . To illustrate the use of this rule, given

$$W = ABC^2 \quad \text{and} \quad X = AB^2,$$

one alias has the form

$$WX = (ABC^2)(AB^2) = A^2B^3C^2 = A^2C^2 = AC,$$

upon reduction of the exponents to the modulus 3. The second alias has the form

$$WX^2 = (ABC^2)(AB^2)^2 = (ABC^2)(A^2B^4) = A^3B^5C^3 = B^2C^2 = BC.$$

It is immaterial which of the components is designated W and which is designated X , since the aliases are also given by the general forms W^2X and WX . As another example, the aliases of the components AB and AC are

$$WX = (AB)(AC) = A^2BC = AB^2C^2,$$

$$WX^2 = (AB)(AC)^2 = A^3BC^2 = BC^2.$$

To construct a design having within-block information on all components of the interactions, and yet provide complete information on all main effects, the scheme presented in Table 8.11-4 may be used. The aliases associated with the pairs of defining relations are indicated. The design represented

Table 8.11-4 Construction of Blocks of Size 3

Replication	Defining relations	Aliases
1	ABC, AB^2	AC^2, BC^2
2	ABC^2, AB^2	AC, BC
3	AB^2C, AB	AC^2, BC
4	AB^2C^2, AB	AC, BC^2

by this scheme will provide some within-block information with respect to all interaction components. For example, the AB^2 component is confounded with block effects in replications 1 and 2 but is free of block effects in replications 3 and 4. Hence the relative within-block information on this component is $\frac{1}{2}$. The AC^2 component is confounded in replications 1 and 3 but is not confounded in blocks 2 and 4. It will be found that all the components of two-factor interactions are confounded with blocks in two of the four replications but are free of such confounding in two of the four replications. Each of the components of the three-factor interaction is confounded with block effects in one of the replications but is free of block effects in three of the four replications. Hence the relative information on the components of the three-factor interaction is $\frac{3}{4}$.

The blocks corresponding to the design outlined in Table 8.11-4 are obtained by subdividing the blocks in Table 8.11-2. Block 1 in the latter table is subdivided as follows:

$(AB^2)_0$	$(AB^2)_1$	$(AB^2)_2$
(000)	(021)	(012)
(111)	(102)	(120)
(222)	(210)	(201)

Each of the treatment combinations within a block satisfies the relation $(ABC)_0$ as well as the relation heading the block. Block 2 in Table 8.11-2 is subdivided as follows:

$(AB^2)_0$	$(AB^2)_1$	$(AB^2)_2$
(001)	(022)	(010)
(112)	(100)	(121)
(220)	(211)	(202)

Block 3 in Table 8.11-2 is subdivided as follows:

$(AB^2)_0$	$(AB^2)_1$	$(AB^2)_2$
(002)	(020)	(011)
(110)	(101)	(122)
(221)	(212)	(200)

The nine blocks of size 3 that have just been constructed form replication 1 of the design in Table 8.11-4.

The over-all analysis of the latter design takes the form given in Table 8.11-5. Within-block information on the main effects is complete; hence no special computational procedures are required. In computing the AB

component, only information from blocks in which this component is not confounded is used. Blocks of this kind are in replications 1 and 2; the blocks in replications 3 and 4 do not provide within-block information on AB . All within-block information on AB^2 is obtained from replications 3 and 4. Hence the sum of squares for AB^2 is computed from a summary table prepared from replications 3 and 4. Given this summary table,

$$SS'_{AB^2} = \frac{(AB^2)'_0^2 + (AB^2)'_1^2 + (AB^2)'_2^2}{18} - \frac{[\Sigma(AB^2)'_i]^2}{54},$$

where $(AB^2)'_i$ is the sum of observations on all treatment combinations in replications 3 and 4 which satisfy the relation $x_1 + 2x_2 = i \pmod{3}$. As a partial check on the numerical work $\Sigma(AB^2)'_i$ is equal to the sum of all observations in replications 3 and 4.

Other components of the two-factor interactions are computed in an analogous manner. The within-block information on AC is obtained from replications 1 and 3; similar information on AC^2 is obtained from replications 2 and 4.

Table 8.11-5 Analysis of Design in Table 8.11-4

Source of variation	df	
Between blocks	35	
Replications	3	
Blocks within reps	32	
Within blocks	72	
Main effects: A, B, C	6	(2 each)
Two-factor int:		
AB, AB^2 ($\frac{1}{2}$ information)	4	(2 each)
AC, AC^2 ($\frac{1}{2}$ information)	4	(2 each)
BC, BC^2 ($\frac{1}{2}$ information)	4	(2 each)
Three-factor int:		
$ABC, ABC^2, AB^2C, AB^2C^2$ ($\frac{3}{4}$ information)	8	(2 each)
Residual	46	

Within-block information on the three-factor interaction components is obtained from some three of the four replications. For example, information from replications 2, 3, and 4 is combined to obtain the ABC component. Given a summary table of observations from these three replications,

$$SS'_{ABC} = \frac{(ABC)'_0^2 + (ABC)'_1^2 + (ABC)'_2^2 + (ABC)'_3^2}{27} - \frac{[\Sigma(ABC)'_i]^2}{81},$$

where $(ABC)'_i$ is the sum of observations in replications 2, 3, and 4 satisfying the relation $x_1 + x_2 + x_3 = i \pmod{3}$. The other components of the three-factor interaction are computed in an analogous manner, the ABC^2 component, for example, being obtained from replications 1, 3, and 4. Since

components of both two-factor and three-factor interactions are partially confounded with block effects, estimates of cell means require adjustment for block effects. The latter adjustment is most readily made by adjusting cell totals. Adjusted cell totals are given by

$$\begin{aligned} ABC'_{ijk} = & \frac{1}{9}(A_i + B_j + C_k) \\ & + \frac{2}{9}[(AB)'_{i+j} + (AB^2)'_{i+2j} + (AC)'_{i+k} + (AC^2)'_{i+2k} + (BC)'_{j+k} + (BC^2)'_{j+2k}] \\ & + \frac{4}{27}[(ABC)'_{i+j+k} + (ABC^2)'_{i+j+2k} + (AB^2C)'_{i+2j+k} + (AB^2C^2)'_{i+2j+2k}] \\ & - \frac{4}{9}G. \end{aligned}$$

The primes indicate that summations are restricted to replications in which a component is not confounded with block effects. The sums required for obtaining an adjusted cell total are obtained in the course of computing the within-block components of the interactions. In terms of the basic linear model, it may be shown that

$$\begin{aligned} ABC'_{ijk} = & 4[\mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + (\alpha\gamma)_{ik} + (\beta\gamma)_{jk} \\ & + (\alpha\beta\gamma)_{ijk}] + 4(\text{sum of all block effects}). \end{aligned}$$

The right-hand side of this last expression represents the parameters that would be estimated by a cell total if each block contained all treatment combinations. Differences between two adjusted totals will be free of block effects, since the last term on the right-hand side is a constant for all cells. With the exception of the block effects, each of the parameters on the right-hand side is estimated independently of all others. Estimates of the parameters associated with main effects are each based upon four effective replications. However, estimates of parameters for two-factor interactions are based upon two effective replications, and estimates of parameters for three-factor interactions are based upon three effective replications. In making comparisons among adjusted cell totals in the three-way table, for most practical purposes one may consider each of these adjusted cell totals as being based upon four replications. Exact methods for taking into account the difference in the effective number of replications for the separate parts of the adjusted cell totals are, however, available [cf. Federer (1955, chap. 9)].

An adjusted total for a cell in a two-way summary table, the cell ab_{ij} , for example, has the form

$$AB'_{ij} = \frac{1}{3}(A_i + B_j) + \frac{2}{3}[(AB)'_{i+j} + (AB^2)'_{i+2j}] - \frac{1}{3}G.$$

Similarly an adjusted total for the cell ac_{ik} has the form

$$AC'_{ik} = \frac{1}{3}(A_i + C_k) + \frac{2}{3}[(AC)'_{i+k} + (AC^2)'_{i+2k}] - \frac{1}{3}G.$$

For comparisons among adjusted totals in a two-way summary table, the effective number of observations in a total varies as a function of the nature of the comparison. The difference between adjusted totals

which are in the same row or column of a two-way summary table is based upon $(\frac{3}{5})(12)$ effective observations, whereas the difference between adjusted totals which are not in the same row or column is based upon $(\frac{2}{3})(12)$ effective observations. Comparisons involving mixtures of these two kinds of differences have effective numbers of observations which may be determined by the general expression for the variance of a linear combination of independent terms.

Table 8.11-6 Analysis of Design in Table 8.11-4 with Repeated Measures

Source of variation	df	
Between subjects	$36n - 1$	
Groups	35	
Subjects within groups	$36(n - 1)$	
Within subjects	$72n$	
Main effects: A, B, C	6 (2 each)	
Two-factor int:		
AB, AB^2 ($\frac{1}{2}$ information)	4	(2 each)
AC, AC^2 ($\frac{1}{2}$ information)	4	(2 each)
BC, BC^2 ($\frac{1}{2}$ information)	4	(2 each)
Three-factor int:		
$ABC, ABC^2, AB^2C, AB^2C^2$ ($\frac{3}{4}$ information)	8	(2 each)
Residual	$72n - 26$	

So far the design outlined in Table 8.11-4 has been considered for the case in which there is one observation per cell. Several variations on this basic design are possible. One variation might be to have the block represent a group of n subjects. Each of the groups would be assigned at random to the blocks, and each of the n subjects within a group would be observed, in a random order, under each of the treatment combinations within a given block. The analysis of the resulting design would have the form shown in Table 8.11-6. Thus, with only three observations per subject, within-subject estimates of all factorial effects are available.

To summarize the principles underlying the construction of designs for $3 \times 3 \times 3$ factorial experiments, blocks of size 9 may be constructed by utilizing relations corresponding to three-factor interactions. If this is done, components of three-factor interactions are either completely or partially confounded with block effects. To construct blocks of size 3, the blocks of size 9 are subdivided by means of relations corresponding to two-factor interactions. In this case, in addition to the interaction corresponding to the defining relations, the aliases also become confounded with block effects.

The principles developed in this section may be extended to include factorial experiments of the form $p \times p \times p$, where p is a prime number. Use

of components of the three-factor interaction reduces the block size to p^2 . Then blocks of size p^2 are reduced to size p through use of relations corresponding to the components of two-factor interactions. For example in a $5 \times 5 \times 5$ factorial experiment, the relations u and v may be used to obtain blocks of size 5. In addition to components corresponding to the relations u and v , components corresponding to the following relations will also be confounded with block effects:

$$u + v, \quad u + 2v, \quad u + 3v, \quad u + 4v.$$

For a single replication with block size 5, the number of blocks required would be 25. Hence the degrees of freedom for blocks within a single replication would be 24. Within a single replication, six components of interactions are confounded with block effects; each component has 4 degrees of freedom. Hence 24 degrees of freedom corresponding to the six components of interactions are confounded with differences between blocks, the latter differences also having 24 degrees of freedom.

In terms of the multiplicative scheme, the aliases of the components W and X in a $5 \times 5 \times 5$ factorial experiment have the general form

$$WX, \quad WX^2, \quad WX^3, \quad WX^4.$$

To make the notation system unique, the exponent of the first letter in an interaction term is made unity by raising the whole term to an appropriate power and then reducing exponents modulo 5. It is not difficult to show that WX^k and $(WX^k)^n$, where n is any number modulo 5, correspond to equivalent relations. For example, the following relations are equivalent in the sense that they define the same set of treatment combinations:

$$AC^2: \quad x_1 + 2x_3 = i \pmod{5},$$

$$A^2C^4: \quad 2x_1 + 4x_3 = 2i \pmod{5}.$$

8.12 Balanced $3 \times 2 \times 2$ Factorial Experiment in Blocks of Size 6

A balanced design for $3 \times 2 \times 2$ factorial experiments in which the block size is 6 can be constructed at the cost of partial confounding of the BC and ABC interactions. A minimum of three replications is required for balance; in this balanced design the relative information on BC is $\frac{8}{9}$, and the relative information on ABC is $\frac{5}{9}$. The balanced design is given in schematic form in Table 8.12-1. In this table the symbol K_0 designates the set of treatment combinations satisfying the relation $x_2 + x_3 = 0 \pmod{2}$, and the symbol K_1 designates the set of treatment combinations satisfying the relation $x_2 + x_3 = 1 \pmod{2}$. In terms of the symbols for individual treatment combinations, the design is given in Table 8.12-2. The spacings within the blocks designate treatment combinations belonging to different K sets.

The computational procedures for all factorial effects, except the BC and ABC interactions, are identical to those for a factorial experiment in which there is no confounding. The BC and ABC interactions require adjustment

for block effects. A simplified procedure for computing the BC interaction will be outlined for the case in which there is no confounding. The adjustments required for the partial confounding with block effects will then be obtained.

Table 8.12-1

Level of A	Rep 1		Rep 2		Rep 3	
	Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
a_0	K_0	K_1	K_1	K_0	K_1	K_0
a_1	K_1	K_0	K_0	K_1	K_1	K_0
a_2	K_1	K_0	K_1	K_0	K_0	K_1

Table 8.12-2

Rep 1		Rep 2		Rep 3	
Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
(000)	(001)	(001)	(000)	(001)	(000)
(011)	(010)	(010)	(011)	(010)	(011)
(101)	(100)	(100)	(101)	(101)	(100)
(110)	(111)	(111)	(110)	(110)	(111)
(201)	(200)	(201)	(200)	(200)	(201)
(210)	(211)	(210)	(211)	(211)	(210)

Assuming that the BC interaction were not confounded with blocks, the total of all observations in the set K_0 , which will be designated by the symbol K_0 , may be shown to be an estimate of the following parameters of the linear model (assuming one observation per cell):

$$K_0 \doteq 18\mu + 18(\beta\gamma)_{K_0}.$$

The symbol $(\beta\gamma)_{K_0}$ denotes the K_0 component of the BC interaction. Similarly, the total of all observations in set K_1 , designated by the symbol K_1 , may be shown to estimate

$$K_1 \doteq 18\mu + 18(\beta\gamma)_{K_1}.$$

Hence the difference between these two totals provides an estimate of

$$K_0 - K_1 \doteq 18(\beta\gamma)_{K_0} - 18(\beta\gamma)_{K_1}.$$

Since the BC interaction has only these two components,

$$(\beta\gamma)_{K_0} + (\beta\gamma)_{K_1} = 0, \quad \text{and hence} \quad (\beta\gamma)_{K_0} = -(\beta\gamma)_{K_1}.$$

(The sum of all components of an interaction term in the general linear model is assumed to be zero, since these terms are measured in deviation units about the grand mean.)

Suppose that the symbol $(\beta\gamma)$ is used to denote $(\beta\gamma)_{K_0}$. Then

$$-(\beta\gamma) = (\beta\gamma)_{K_1}.$$

Hence

$$K_0. - K_{1.} \doteq 18(\beta\gamma) - [-18(\beta\gamma)] = 36(\beta\gamma).$$

Thus, if there were no confounding with blocks, the difference between the sum of all observations in set K_0 and all observations in set K_1 estimates a parameter which is equal in absolute value to the difference between the K components of the BC interaction. Since the BC interaction has only one degree of freedom, there is only one independent parameter associated with it. In terms of the parameters of the model, disregarding the experimental error, the sum of squares for the BC interaction would be

$$SS_{BC} \doteq 18(\beta\gamma)_{K_0}^2 + 18(\beta\gamma)_{K_1}^2 = 36(\beta\gamma)^2.$$

Thus a computational formula for the BC interaction, if there were no confounding with blocks, would be

$$SS_{BC} = \frac{(K_0. - K_{1.})^2}{36} \doteq 36(\beta\gamma)^2.$$

In the design in Table 8.12-2, however, the BC interaction is partially confounded with block effects. The difference between the total of all observations in the set K_0 and all observations in the set K_1 actually estimates a mixture of $(\beta\gamma)$ and block effects, i.e.,

$$K_0. - K_{1.} \doteq 36(\beta\gamma) + 2(\pi_2 + \pi_4 + \pi_6 - \pi_1 - \pi_3 - \pi_5),$$

where the π 's represent block effects. If P_i represents the total of all observations in block i ,

$$\left(\frac{1}{3}\right)(P_2 + P_4 + P_6 - P_1 - P_3 - P_5) \doteq 4(\beta\gamma) + 2(\pi_2 + \pi_4 + \pi_6 - \pi_1 - \pi_3 - \pi_5).$$

The quantity on the left-hand side of this last equation may be used to adjust the difference $K_0. - K_{1.}$ for block effects. The adjusted difference takes the form

$$\begin{aligned} Q &= K_0. - K_{1.} - \left(\frac{1}{3}\right)(P_2 + P_4 + P_6 - P_1 - P_3 - P_5) \\ &\doteq 32(\beta\gamma). \end{aligned}$$

Thus the adjusted difference between the K totals provides 32 effective observations in the estimation of $(\beta\gamma)$; but if there were no confounding with blocks, the unadjusted difference between the K totals would provide 36 effective observations. Hence the relative effective information on BC is $\frac{32}{36}$, or $\frac{8}{9}$. The adjusted sum of squares for BC takes the form

$$SS'_{BC} = \frac{Q^2}{32} \doteq 32(\beta\gamma)^2.$$

The problem of finding the appropriate adjustment for the ABC interaction follows the same general pattern as that which has just been indicated

for the BC interaction. A simplified procedure for finding the ABC interaction will first be outlined for the case in which there is no confounding. The symbol K_{0i} will be used to designate the subset of all treatment combinations in the set K_0 which are at level a_i ; similarly the symbol K_{1i} will be used to designate the subset of treatment combinations in the set K_1 which are at level a_i . The totals for all observations in the respective subsets will be denoted by the symbols K_{0i} and K_{1i} . If the symbol $(\alpha\beta\gamma)_{K_{0i}}$ denotes an effect associated with the ABC interaction,

$$K_{0i} - K_{1i} \doteq 6(\alpha\beta\gamma)_{K_{0i}} - 6(\alpha\beta\gamma)_{K_{1i}} + 12(\beta\gamma).$$

Since the sum of the components of an interaction at a fixed level of one of the factors is zero,

$$(\alpha\beta\gamma)_{K_{0i}} + (\alpha\beta\gamma)_{K_{1i}} = 0; \quad \text{hence} \quad (\alpha\beta\gamma)_{K_{0i}} = -(\alpha\beta\gamma)_{K_{1i}}.$$

If the symbol $(\alpha\beta\gamma)_i$ designates either $(\alpha\beta\gamma)_{K_{0i}}$ or $-(\alpha\beta\gamma)_{K_{1i}}$, assuming there were no confounding with blocks,

$$K_{0i} - K_{1i} \doteq 12(\alpha\beta\gamma)_i + 12(\beta\gamma).$$

Specifically, for each level of factor A ,

$$K_{00} - K_{10} \doteq 12(\alpha\beta\gamma)_0 + 12(\beta\gamma),$$

$$K_{01} - K_{11} \doteq 12(\alpha\beta\gamma)_1 + 12(\beta\gamma),$$

$$K_{02} - K_{12} \doteq 12(\alpha\beta\gamma)_2 + 12(\beta\gamma).$$

Thus, if no confounding were present,

$$\frac{E(K_{0i} - K_{1i})^2}{12} \doteq 12(\alpha\beta\gamma)_0^2 + 12(\alpha\beta\gamma)_1^2 + 12(\alpha\beta\gamma)_2^2 + 36(\beta\gamma)^2.$$

The sum of the first three terms on the right-hand side of the last equation defines SS_{ABC} in terms of the parameters of the model; the last term defines SS_{BC} . (In both cases the error component has not been included.) Hence

$$\frac{\Sigma(K_{0i} - K_{1i})^2}{12} = SS_{ABC} + SS_{BC}.$$

Thus, if there were no confounding, a computational formula for SS_{ABC} would be given by

$$SS_{ABC} = \frac{\Sigma(K_{0i} - K_{1i})^2}{12} - \frac{(K_0 - K_1)^2}{36}.$$

In words, the ABC sum of squares is obtained by summing the BC interaction at each level of factor A and then subtracting the over-all BC interaction.

$$SS_{ABC} = \Sigma SS_{BC \text{ at } a_i} - SS_{BC}.$$

For factorial experiments in which some of the factors are at two levels, this general computational procedure is simpler than direct calculation of the three-factor interaction.

In the design under consideration the difference $K_{00} - K_{10}$ is not free of block effects. For this design

$$K_{00} - K_{10} \doteq 12(\alpha\beta\gamma)_0 + 12(\beta\gamma) + 2(\pi_1 + \pi_4 + \pi_6 - \pi_2 - \pi_3 - \pi_5).$$

The blocks in which K_0 appears at level a_0 have positive signs; the blocks in which K_1 appears at level a_0 have negative signs. In obtaining the adjustment for block effects it is more convenient to work with the expression

$$3(K_{00} - K_{10}) \doteq 36(\alpha\beta\gamma)_0 + 36(\beta\gamma) + 6(\pi_1 + \pi_4 + \pi_6 - \pi_2 - \pi_3 - \pi_5).$$

The adjustment for block effects requires the term

$$(\text{adj } a_0) = P_1 + P_4 + P_6 - P_2 - P_3 - P_5 \doteq 16(\alpha\beta\gamma)_0 + 4(\beta\gamma) + 6(\pi_1 + \pi_4 + \pi_6 - \pi_2 - \pi_3 - \pi_5).$$

The adjusted difference used to obtain the sum of squares for the three-factor interaction is

$$3R_0 = 3(K_{00} - K_{10}) - (\text{adj } a_0) - Q \doteq 20(\alpha\beta\gamma)_0,$$

where Q is the quantity used in the computation of SS'_{BC} . Other adjustments for the levels of factor A are

$$(\text{adj } a_1) = P_2 + P_3 + P_6 - P_1 - P_4 - P_5,$$

$$(\text{adj } a_2) = P_2 + P_4 + P_5 - P_1 - P_3 - P_6.$$

The adjusted differences used in the computation of SS'_{ABC} are

$$3R_1 = 3(K_{01} - K_{11}) - (\text{adj } a_1) - Q \doteq 20(\alpha\beta\gamma)_1,$$

$$3R_2 = 3(K_{02} - K_{12}) - (\text{adj } a_2) - Q \doteq 20(\alpha\beta\gamma)_2.$$

An adjusted difference of the form $3R_i$ provides 20 effective observations in the estimation of the parameter $(\alpha\beta\gamma)_i$, whereas the corresponding unadjusted difference provides 36 effective observations. Hence the relative information for the ABC interaction in this design is $\frac{20}{36} = \frac{5}{9}$. A computational formula for the adjusted sum of squares for the ABC interaction is given by

$$SS'_{ABC} = \frac{\Sigma(3R_i)^2}{9(\frac{20}{3})} \doteq (\frac{20}{3})[(\alpha\beta\gamma)_0^2 + (\alpha\beta\gamma)_1^2 + (\alpha\beta\gamma)_2^2].$$

If there were no confounding with blocks,

$$SS_{ABC} \doteq 12[(\alpha\beta\gamma)_0^2 + (\alpha\beta\gamma)_1^2 + (\alpha\beta\gamma)_2^2],$$

where the error component has been disregarded. The ratio $(\frac{20}{3})/12 = \frac{5}{9}$, the relative information for ABC . The over-all analysis of the balanced design in Table 8.12-2 is given in Table 8.12-3. In the underlying model, all factors are assumed to be fixed, blocks are assumed to be random, and the block \times treatment interactions are assumed to be zero or negligible. Only under these stringent assumptions are the adjustments for block effects valid. In making F tests on all factorial effects, MS_{res} forms the denominator of all F ratios. In making comparisons between adjusted cell means,

Table 8.12-3

Source of variation	SS	df	MS	E(MS)
Between blocks		5		
Replications		2		
Blocks within reps		3		
Within blocks		30		
<i>A</i>		2		$\sigma_e^2 + nqr\sigma_\alpha^2$
<i>B</i>		1		$\sigma_e^2 + npr\sigma_\beta^2$
<i>C</i>		1		$\sigma_e^2 + npq\sigma_\gamma^2$
<i>AB</i>		2		$\sigma_e^2 + nr\sigma_{\alpha\beta}^2$
<i>AC</i>		2		$\sigma_e^2 + nq\sigma_{\alpha\gamma}^2$
<i>(BC)'</i>		1		$\sigma_e^2 + (\frac{8}{9})np\sigma_{\beta\gamma}^2$
<i>(ABC)'</i>		2		$\sigma_e^2 + (\frac{8}{9})n\sigma_{\alpha\beta\gamma}^2$
Residual		19		σ_e^2

the effective number of observations on the component parts should be taken into account.

The adjustment for \overline{BC}_{00} is obtained by considering the parameters actually estimated by the total of all observations under treatment combination bc_{00} , the levels of factor *A* being disregarded. There are nine such observations.

$$BC_{00} \doteq 9\mu + 9\beta_0 + 9\gamma_0 + 9(\beta\gamma)_{00} + (\pi_1 + \pi_3 + \pi_5) + 2(\pi_2 + \pi_4 + \pi_6).$$

To obtain an adjustment for BC_{00} , one must construct an expression which estimates the block effects on the right-hand side of this last equation. This expression is given by

$$\begin{aligned} (\text{adj } BC_{00}) &= (\frac{1}{6})(P_1 + P_3 + P_5) + (\frac{1}{3})(P_2 + P_4 + P_6) - (\frac{1}{3})Q - (\frac{1}{4})G \\ &\doteq (\pi_1 + \pi_3 + \pi_5) + 2(\pi_2 + \pi_4 + \pi_6). \end{aligned}$$

The adjusted total for the cell bc_{00} is thus

$$BC'_{00} = BC_{00} - (\text{adj } BC_{00}) \doteq 9\mu + 9\beta_0 + 9\gamma_0 + 9(\beta\gamma)_{00}.$$

The adjusted mean for cell bc_{00} is given by

$$\overline{BC}_{00} = \frac{BC'_{00}}{9} \doteq \mu + \beta_0 + \gamma_0 + (\beta\gamma)_{00}.$$

Since the treatment combination bc_{00} and bc_{11} both belong to the set K_0 , and since the design is symmetrical with respect to the parts of K_0 , the adjustment for \overline{BC}_{11} is the same as that used for \overline{BC}_{00} .

By similar arguments and use of the relation that $(\beta\gamma)_{K_0} = -(\beta\gamma)_{K_1}$, the adjustment for both \overline{BC}_{01} and \overline{BC}_{10} may be shown to be

$$(\text{adj } BC_{01}) = (\text{adj } BC_{10}) = \left(\frac{1}{3}\right)(P_1 + P_3 + P_5) + \left(\frac{1}{6}\right)(P_2 + P_4 + P_6) + \left(\frac{1}{32}\right)Q - \left(\frac{1}{4}\right)G.$$

In the arguments that have been used in the course of arriving at various adjustments, certain assumptions about restrictions on the parameters in the general linear model have been invoked. These restrictions are made explicit in Table 8.12-4.

Table 8.12-4

	c_0		c_1				
	b_0	b_1	b_0	b_1		b_0	b_1
a_0	$(\alpha\beta\gamma)_0$	$-(\alpha\beta\gamma)_0$	$-(\alpha\beta\gamma)_0$	$(\alpha\beta\gamma)_0$	c_0	$(\beta\gamma)$	$-(\beta\gamma)$
a_1	$(\alpha\beta\gamma)_1$	$-(\alpha\beta\gamma)_1$	$-(\alpha\beta\gamma)_1$	$(\alpha\beta\gamma)_1$	c_1	$-(\beta\gamma)$	$(\beta\gamma)$
a_2	$(\alpha\beta\gamma)_2$	$-(\alpha\beta\gamma)_2$	$-(\alpha\beta\gamma)_2$	$(\alpha\beta\gamma)_2$			
Sum	0	0	0	0	Sum	0	0

8.13 Numerical Example of $3 \times 2 \times 2$ Factorial Experiment in Blocks of Size 6

Data for this type of design will generally take the form given in Table 8.13-1. The observation is given opposite the symbol for the treatment

Table 8.13-1

Rep 1		Rep 2		Rep 3	
Block 1	Block 2	Block 3	Block 4	Block 5	Block 6
(000) 4	(001) 10	(001) 5	(000) 3	(001) 5	(000) 3
(011) 5	(010) 20	(010) 10	(011) 4	(010) 10	(011) 1
(101) 15	(100) 5	(100) 15	(101) 5	(101) 10	(100) 0
(110) 20	(111) 10	(111) 15	(110) 5	(110) 5	(111) 5
(201) 20	(200) 10	(201) 10	(200) 10	(200) 10	(201) 20
(210) 10	(211) 1	(210) 5	(211) 5	(211) 4	(210) 5
$P_1 = 74$	$P_2 = 56$	$P_3 = 60$	$P_4 = 32$	$P_5 = 44$	$P_6 = 34$

$G = 300$

combination. The make-up of the blocks is that given in Table 8.12-1. In the behavioral sciences the blocks may correspond to an individual or to a group of individuals.

Summary data for the factorial effects are given in Table 8.13-2. From part iv of this table one obtains K_0 and K_1 ,

$$K_0 = 60 + 50 = 110,$$

$$K_1 = 90 + 100 = 190.$$

Table 8.13-2

(i)

		b_0		b_1		
		c_0	c_1	c_0	c_1	
a_0		10	20	40	10	80
a_1		20	30	30	30	110
a_2		30	50	20	10	110
		60	100	90	50	300

(ii)

	b_0	b_1
a_0	30	50
a_1	50	60
a_2	80	30
	160	140

(iii)

	c_0	c_1
a_0	50	30
a_1	50	60
a_2	50	60
	150	150

(iv)

	c_0	c_1
b_0	60	100
b_1	90	50
	150	150

From the block totals in Table 8.13-1, one finds

$$P_2 + P_4 + P_6 - P_1 - P_3 - P_5 = -56.$$

Hence

$$Q = K_0 - K_1 - \frac{P_2 + P_4 + P_6 - P_1 - P_3 - P_5}{3}$$

$$= 110 - 190 - \frac{(-56)}{3}$$

$$= -61.33.$$

Having the numerical value of Q , one obtains the adjusted sum of squares for BC from the relation

$$SS'_{BC} = \frac{Q^2}{32} = \frac{(-61.33)^2}{32} = 117.54.$$

Expressions of the form $3(K_{0i} - K_{1i})$ are obtained from part i of Table 8.13-2.

$$3(K_{00} - K_{10}) = 3[(10 + 10) - (20 + 40)] = -120,$$

$$3(K_{01} - K_{11}) = 3[(20 + 30) - (30 + 30)] = -30,$$

$$3(K_{02} - K_{12}) = 3[(30 + 10) - (50 + 20)] = -90.$$

Adjustments for the above expressions are respectively,

$$(\text{adj } a_0) = P_1 + P_4 + P_6 - P_2 - P_3 - P_5 = -20$$

$$(\text{adj } a_1) = P_2 + P_3 + P_6 - P_1 - P_4 - P_5 = 0$$

$$(\text{adj } a_2) = \frac{P_2 + P_4 + P_5 - P_1 - P_3 - P_6}{2} = \frac{-36}{2}$$

$$\text{Sum} = P_2 + P_4 + P_6 - P_1 - P_3 - P_5 = -56$$

The R 's required for computing SS'_{ABC} are given by

$$3R_0 = 3(K_{00.} - K_{10.}) - (\text{adj } a_0) - Q = -38.67$$

$$3R_1 = 3(K_{01.} - K_{11.}) - (\text{adj } a_1) - Q = 31.33$$

$$3R_2 = 3(K_{02.} - K_{12.}) - (\text{adj } a_2) - Q = 7.33$$

$$\Sigma(3R) = 3(-80) - (-56) - 3(-61.33) = -.01$$

Within rounding error, $\Sigma(3R) = 0$.

Table 8.13-3

(1) = $G^2/36$	= 2500	(7) = $[\Sigma(AC)^2]/6$	= 2600.00
(2) = ΣX^2	= 3618	(8) = $[\Sigma(BC)^2]/9$	= 2688.89
(3) = $(\Sigma A^2)/12$	= 2550.00	(9) = $[\Sigma(ABC)^2]/3$	= 3066.67
(4) = $(\Sigma B^2)/18$	= 2511.11	(10) = $(\Sigma P^2)/6$	= 2721.33
(5) = $(\Sigma C^2)/18$	= 2500.00	(11) = $(\Sigma \text{Rep}^2)/12$	= 2620.67
(6) = $[\Sigma(AB)^2]/6$	= 2800.00		
SS_{blocks}	= (10) - (1)		= 221.33
SS_{reps}	= (11) - (1)		= 120.67
$SS_{\text{blocks w. rep}}$	= (10) - (11)		= 100.66
$SS_{\text{w. block}}$	= (2) - (10)		= 896.67
SS_A	= (3) - (1)		= 50.00
SS_B	= (4) - (1)		= 11.11
SS_C	= (5) - (1)		= 0
SS_{AB}	= (6) - (3) - (4) + (1)		= 238.89
SS_{AC}	= (7) - (3) - (5) + (1)		= 50.00
$SS_{BC}(\text{unadj})$	= (8) - (4) - (5) + (1)		= 177.78
$SS_{BC}(\text{adj})$	= SS'_{BC}		= (117.54)
$SS_{ABC}(\text{unadj})$	= (9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)		= 38.89
$SS_{ABC}(\text{adj})$	= SS'_{ABC}		= (42.18)
$SS_{\text{res}}(\text{unadj})$	= (2) - (9) - (10) + (1)		= 330.00
$SS_{\text{res}}(\text{adj})$	= $SS_{\text{res}}(\text{unadj}) + SS_{BC}(\text{unadj}) + SS_{ABC}(\text{unadj})$ - $SS_{BC}(\text{adj}) - SS_{ABC}(\text{adj})$		= 386.95

The adjusted sum of squares for ABC is

$$SS'_{ABC} = \frac{\Sigma(3R_i)^2}{60} = 42.18.$$

A summary of the computation of all sums of squares required in the analysis of variance is given in Table 8.13-3. The residual term in this analysis may be obtained either by subtracting the total of the adjusted treatment sums of squares from $SS_{\text{w. block}}$ or by the method indicated in the table.

The relative efficiency of a design involving small block size as compared with a completely randomized design depends in large part upon the relative magnitudes of $MS_{\text{blocks w. rep}}$ and $MS_{\text{res}}(\text{adj})$. For the data under consideration these mean squares are, respectively, 33.55 and 20.37. The smaller the latter mean square relative to the former, the more efficient the design involving the smaller block size.

To illustrate the computation of the adjustments for the BC means,

$$\begin{aligned}(\text{adj } BC_{00}) &= (\tfrac{1}{6})(P_1 + P_3 + P_5) + (\tfrac{1}{3})(P_2 + P_4 + P_6) - (\tfrac{1}{32})Q - (\tfrac{1}{4})G \\&= (\tfrac{1}{6})(178) + (\tfrac{1}{3})(122) - (-1.92) - 75 \\&= 29.67 + 40.67 + 1.92 - 75 = -2.74.\end{aligned}$$

The adjusted mean is

$$\overline{BC}'_{00} = \overline{BC}_{00} - \frac{(-2.74)}{9} = 6.67 - \frac{(-2.74)}{9} = 6.97.$$

The adjustment for \overline{BC}_{01} will be numerically equal but opposite in sign to the adjustment for \overline{BC}_{00} . Thus

$$\overline{BC}'_{01} = \overline{BC}_{01} - \frac{2.74}{9} = 11.11 - \frac{2.74}{9} = 10.81.$$

The restrictions on the parameters underlying the linear model for a $3 \times 2 \times 2$ factorial experiment as shown in Table 8.12-3 may be illustrated numerically through use of the data in Table 8.13-2. These restrictions also hold for estimates of these parameters. For example, assuming no confounding,

$$\begin{aligned}(\alpha\beta\gamma)_{000} &\doteq \frac{ABC_{000}}{3} - \frac{AB_{00}}{6} - \frac{AC_{00}}{6} - \frac{BC_{00}}{9} + \frac{A_0}{12} + \frac{B_0}{18} + \frac{C_0}{18} - \frac{G}{36} \\&\doteq -1.11.\end{aligned}$$

According to Table 8.12-3, $(\alpha\beta\gamma)_{000} = -(\alpha\beta\gamma)_{000}$. The latter parameter is estimated by

$$\begin{aligned}(\alpha\beta\gamma)_{001} &\doteq \frac{ABC_{001}}{3} - \frac{AB_{00}}{6} - \frac{AC_{01}}{6} - \frac{BC_{01}}{9} + \frac{A_0}{12} + \frac{B_0}{18} + \frac{C_0}{18} - \frac{G}{36} \\&\doteq 1.11.\end{aligned}$$

Thus $(\alpha\beta\gamma)_{000} = -(\alpha\beta\gamma)_{001} = (\alpha\beta\gamma)_0 \doteq -1.11$.

8.14 $3 \times 3 \times 3 \times 2$ Factorial Experiment in Blocks of Size 6

The principles used in the construction of the $3 \times 2 \times 2$ factorial design for block size 6 may be extended to cover a $3 \times 3 \times 3 \times 2$ factorial experiment with block size 6. To illustrate the construction of this design, the following notation will be used:

L_{ij} = set of all treatment combinations satisfying both the relations

$$x_1 + x_2 = i \pmod{3} \quad \text{and} \quad x_1 + x_3 = j \pmod{3}.$$

For example, the fourth factor being disregarded, the treatment combinations in the set L_{00} have the form $(000-)$, $(122-)$, and $(211-)$. As another example, the treatment combinations in the set L_{12} have the form $(012-)$, $(101-)$, and $(220-)$. For each of the treatment combinations in the set L_{12} the relations $x_1 + x_2 = 1 \pmod{3}$ and $x_1 + x_3 = 2 \pmod{3}$ are satisfied.

Each of the sets L_{ij} consists of three treatment combinations at level d_0 and three at level d_1 .

The design in Table 8.14-1 represents a partially balanced design. The actual treatment combinations in this design are given in Table 8.14-2. A

Table 8.14-1

Level of D	Block:	1	2	3	4	5	6	7	8	9
d_0		L_{00}	L_{01}	L_{02}	L_{10}	L_{11}	L_{12}	L_{20}	L_{21}	L_{22}
d_1		L_{12}	L_{22}	L_{11}	L_{01}	L_{20}	L_{21}	L_{02}	L_{00}	L_{10}

Table 8.14-2

Block:	1	2	3	4	5	6	7	8	9
	(0000)	(0010)	(0020)	(0100)	(0110)	(0120)	(0200)	(0210)	(0220)
	(1220)	(1200)	(1210)	(1020)	(1000)	(1010)	(1120)	(1100)	(1110)
	(2110)	(2120)	(2100)	(2210)	(2220)	(2200)	(2010)	(2020)	(2000)
	(0121)	(0221)	(0111)	(0011)	(0201)	(0211)	(0021)	(0001)	(0101)
	(1011)	(1111)	(1001)	(1201)	(1121)	(1101)	(1211)	(1221)	(1021)
	(2201)	(2001)	(2221)	(2121)	(2011)	(2021)	(2101)	(2111)	(2211)

minimum of four replications is required for complete balance. To construct the latter, the defining relations for i and j may be changed; i.e., the relations

$$x_1 + x_2 = i \pmod{3}, \quad x_1 + 2x_3 = j \pmod{3}$$

will define sets of L 's which could make up a second replication. The assignment of the L 's defined in this manner would be identical to that shown in Table 8.14-1.

The design given in this table has the following restrictions imposed upon the assignment of the L 's to the blocks: (1) All the possible L 's occur once and only once at each level of factor D . (2) Within each block a subscript does not occur twice in the same position; that is, L_{01} and L_{02} cannot occur within the same block, since the subscript zero would be repeated within the same position within the same block. (3) The sum of the subscripts (modulo 3) for L 's within the same block must be equal. For example, in block 9 one finds L_{22} and L_{10} in the same block; the sum of the subscripts for L_{22} is $2 + 2 = 1 \pmod{3}$, and the sum of the subscripts for L_{10} is $1 + 0 = 1 \pmod{3}$.

In the original definition of L_{ij} , the relations used correspond to the AB and AC components of the $A \times B$ and $A \times C$ interactions, respectively. The relative information on these components is $\frac{3}{4}$. The generalized interactions of the AB and AC components are BC^2 and AB^2C^2 . The relative information on BC^2 is $\frac{3}{4}$, but there is no within-block information on the

AB^2C^2 component. The ABC , ACD , and B^2C^2D components are also partially confounded with block effects; the relative within-block information on each of these components is $\frac{1}{4}$. However, the AB^2C^2D component is not confounded with block effects.

To show that the design in Table 8.14-1 provides no within-block information on the AB^2C^2 component, it is necessary to show that no comparisons which belong to this component can be obtained from information provided by a single block. The treatment combinations defining the AB^2C^2 sets are located in the following blocks:

	Blocks
$x_1 + 2x_2 + 2x_3 = 0 \pmod{3}$	1, 6, 8
$x_1 + 2x_2 + 2x_3 = 1 \pmod{3}$	3, 5, 7
$x_1 + 2x_2 + 2x_3 = 2 \pmod{3}$	2, 4, 9

Thus no single block contains treatment combinations which belong to more than one of the AB^2C^2 sets. Hence any comparison between such sets will involve differences between blocks.

The picture with respect to the AB^2C^2 component may be contrasted with that presented by the AB component, for which there is $\frac{3}{4}$ relative within-block information. The blocks which contain treatment combinations belonging to the AB sets are the following:

	Blocks
$J_0: x_1 + x_2 = 0 \pmod{3}$	1, 2, 3, 4, 7, 8
$J_1: x_1 + x_2 = 1 \pmod{3}$	1, 3, 4, 5, 6, 9
$J_2: x_1 + x_2 = 2 \pmod{3}$	2, 5, 6, 7, 8, 9

Thus within-block information on the AB component is available on the following comparisons in the blocks indicated:

	Blocks
J_0 versus J_1	1, 3, 4
J_0 versus J_2	2, 7, 8
J_1 versus J_2	5, 6, 9

In the design given in Table 8.14-1 each J set occurs in the same block with each of the other J sets three times.

Procedures for obtaining the adjusted sums squares for the two-factor interactions which are partially confounded with block effects will be illustrated by working with the AB (or J) component. The symbol J_0 will be used to represent the set of treatment combinations which satisfy the relation $x_1 + x_2 = 0 \pmod{3}$. For the design in Table 8.14-1, these treatment combinations occur in cells of the form L_{0j} , that is, cells in which the first subscript of an L is zero. The symbol J_0 will be used to designate the sum of all treatment combinations in the set J_0 . In terms of the parameters of the general linear model,

$$2J_0 = 2\sum_j L_{0j} = 36\mu + 36(\alpha\beta)_{J_0} + 6(\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_7 + \pi_8).$$

In words, the right-hand side of the last expression indicates that the total J_0 is partially confounded with block effects. The symbol $\Sigma P_{L_{0j}}$ will be used to designate the sum of totals for blocks in which treatment combinations belonging to J_0 are located. Thus

$$\Sigma P_{L_{0j}} = P_1 + P_2 + P_3 + P_4 + P_7 + P_8.$$

In terms of the parameters of the general linear model,

$$\Sigma P_{L_{0j}} \doteq 36\mu + 9(\alpha\beta)_{J_0} + 6(\pi_1 + \pi_2 + \pi_3 + \pi_4 + \pi_7 + \pi_8).$$

A quantity convenient for use in obtaining the adjusted sum of squares for the J component of $A \times B$ is

$$2Q_{J_0} = 2J_0 - \Sigma P_{L_{0j}} \doteq 27(\alpha\beta)_{J_0}.$$

The coefficient of $(\alpha\beta)_{J_0}$ is 36 in the $2J_0$ total and 27 in the $2Q_{J_0}$ total. The ratio $\frac{27}{36} = \frac{3}{4}$ gives the relative within-block information for the AB component.

The other quantities required for the computation of the adjusted sum of squares for the AB component are

$$2Q_{J_1} = 2J_1 - \Sigma P_{L_{1j}} \doteq 27(\alpha\beta)_{J_1},$$

where $P_{L_{1j}}$ represents a total for a block containing treatment combinations belonging to the set J_1 , and

$$2Q_{J_2} = 2J_2 - \Sigma P_{L_{2j}} \doteq 27(\alpha\beta)_{J_2}.$$

The adjusted sum of squares for the AB component is given by

$$SS'_{J(AB)} = \frac{\Sigma(2Q_{J_i})^2}{108} = (\frac{27}{2})\Sigma(\alpha\beta)_{J_i}^2.$$

(The error component has been disregarded in this last expression.) If the AB component were not partially confounded with block effects, the sum of squares for this component would be given by

$$SS_{J(AB)} = \frac{\Sigma(2J_i)^2}{144} - \frac{[\Sigma(2J_i)]^2}{432} \doteq 18\Sigma(\alpha\beta)_{J_i}^2.$$

[The ratio of the coefficients of $\Sigma(\alpha\beta)^2$ provides the formal definition of relative within-block information. In this case the ratio is $(\frac{27}{2})/18 = \frac{3}{4}$.]

Adjustments for the AC component of $A \times C$ are obtained in an analogous manner. If K_j designates the set of treatment combinations which satisfy the relation $x_1 + x_3 = j \pmod{3}$, and if K_j designates the sum of all observations on treatment combinations belonging to the set K_j , then

$$2Q_{K_j} = 2K_j - \sum_i P_{L_{ij}} \doteq 27(\alpha\gamma)_{K_j}.$$

For example,

$$2Q_{K_0} = 2K_0 - \sum_i P_{L_{i0}} \doteq 27(\alpha\gamma)_{K_0}.$$

The adjusted sum of squares for the AC component is

$$SS'_{K(AC)} = \frac{\Sigma(2Q_{Kj})^2}{108} \doteq (\frac{27}{2})\Sigma(\alpha\gamma)_{Kj}^2.$$

Adjustments for the BC^2 component may be cast in the following form:

$$2Q_{(BC^2)_m} = 2(BC^2)_m + \Sigma P_{(BC^2)_m} \doteq 27(\beta\gamma^2)_m.$$

In this expression $(BC^2)_m$ designates the sum of all observations on treatment combinations which satisfy the relation $x_2 + 2x_3 = m \pmod{3}$, and $P_{(BC^2)_m}$ designates a total for a block containing treatment combinations which belong to the set $(BC^2)_m$. Blocks which have treatment combinations in the latter set are those containing L_{ij} 's which satisfy the relation $i + 2j = m \pmod{3}$. For example, treatment combinations in the set $(BC^2)_0$ are included in L_{00} , L_{22} , and L_{11} ; these L 's are located in blocks 1, 2, 3, 5, 8, and 9. Hence

$$\Sigma P_{(BC^2)_0} = P_1 + P_2 + P_3 + P_5 + P_8 + P_9.$$

The process of obtaining the adjustments for the three-factor interactions which are partially confounded with block effects is simplified by utilizing the following restrictions on the underlying parameters,

$$(\alpha\beta\delta)_{J_0 \text{ at } d_0} + (\alpha\beta\delta)_{J_0 \text{ at } d_1} = 0,$$

and hence

$$(\alpha\beta\delta)_{J_0 \text{ at } d_0} = -(\alpha\beta\delta)_{J_0 \text{ at } d_1} = (\alpha\beta\delta)_{J_0}.$$

Analogous restrictions hold for J_1 and J_2 .

$$(\alpha\beta\delta)_{J_1 \text{ at } d_0} = -(\alpha\beta\delta)_{J_1 \text{ at } d_1} = (\alpha\beta\delta)_{J_1},$$

$$(\alpha\beta\delta)_{J_2 \text{ at } d_0} = -(\alpha\beta\delta)_{J_2 \text{ at } d_1} = (\alpha\beta\delta)_{J_2}.$$

If there were no confounding with blocks, the difference between the totals $J_{0. \text{ at } d_0}$ and $J_{0. \text{ at } d_1}$ would be a function only of the term $(\alpha\beta\gamma)_{J_0}$. This result follows from the basic definition of a three-factor interaction; in this case the ABD interaction is a measure of the difference between AB profiles at the two levels of factor D . Since the AB component is partially confounded with block effects, the $AB \times D$ component will also be partially confounded with block effects.

$$2J_{0. \text{ at } d_0} - 2J_{0. \text{ at } d_1} \doteq 36(\alpha\beta\delta)_{J_0} + 6(\pi_1 + \pi_2 + \pi_3 - \pi_4 - \pi_7 - \pi_8).$$

The last term in the above expression represents the block effects with which the three-factor interaction is confounded. Blocks containing L_{0j} at level d_0 appear with positive signs; blocks containing L_{0j} at level d_1 appear with negative signs. To adjust for block effects,

$$\Sigma P_{L_{0j} \text{ at } d_0} - \Sigma P_{L_{0j} \text{ at level } d_1} \doteq 27(\alpha\beta\delta)_{J_0} + 6(\pi_1 + \pi_2 + \pi_3 - \pi_4 - \pi_7 - \pi_8).$$

The expression $\Sigma P_{L_{0j} \text{ at } d_0}$ denotes the sum of totals for all blocks which contain L_{0j} at level d_0 . The right-hand side of the above expression makes

use of the fact that $(\alpha\beta\delta)_{J_0} + (\alpha\beta\delta)_{J_1} + (\alpha\beta\delta)_{J_2} = 0$. An estimate of $(\alpha\beta\delta)_{J_0}$, which is free of block effects, is given by

$$\begin{aligned} 2R_{J_0} &= 2J_{0, \text{at } d_0} - 2J_{0, \text{at } d_1} - (\Sigma P_{L_{0j} \text{ at } d_0} - \Sigma P_{L_{0j} \text{ at } d_1}) \\ &\doteq 9(\alpha\beta\delta)_{J_0}. \end{aligned}$$

If there were no confounding with blocks, the coefficient of $(\alpha\beta\delta)_{J_0}$ would be 36. The ratio $\frac{9}{36} = \frac{1}{4}$ gives the relative within-block information on the $AB \times D$ component of the $A \times B \times D$ interaction. Other quantities required in order to compute $SS'_{AB \times D}$ are

$$\begin{aligned} 2R_{J_1} &= 2J_{1, \text{at } d_0} - 2J_{1, \text{at } d_1} - (\Sigma P_{L_{1j} \text{ at } d_0} + \Sigma P_{L_{1j} \text{ at } d_1}) \\ &\doteq 9(\alpha\beta\delta)_{J_1}, \\ 2R_{J_2} &= 2J_{2, \text{at } d_0} - 2J_{2, \text{at } d_1} - (\Sigma P_{L_{2j} \text{ at } d_0} + \Sigma P_{L_{2j} \text{ at } d_1}) \\ &\doteq 9(\alpha\beta\delta)_{J_2}. \end{aligned}$$

The adjusted sum of squares for $AB \times D$ is

$$SS'_{AB \times D} = \frac{\Sigma 2R_{J_i}^2}{36} \doteq 9\Sigma(\alpha\beta\delta)_{J_i}^2.$$

Quantities required for the adjusted sum of squares for the $AC \times D$ component have the form

$$\begin{aligned} 2R_{K_j} &= 2K_{j, \text{at } d_0} - 2K_{j, \text{at } d_1} - \left(\Sigma_i P_{L_{ij} \text{ at } d_0} + \Sigma_i P_{L_{ij} \text{ at } d_1} \right) \\ &\doteq 9(\alpha\gamma\delta)_{K_j}. \end{aligned}$$

The corresponding quantities for the $BC^2 \times D$ component have the form

$$\begin{aligned} 2R_{(BC^2)_m} &= (BC^2)_{m \text{ at } d_0} - (BC^2)_{m \text{ at } d_1} - (\Sigma P_{(BC^2)_m \text{ at } d_0} + \Sigma P_{(BC^2)_m \text{ at } d_1}) \\ &\doteq 9(\beta\gamma^2\delta)_m. \end{aligned}$$

8.15 Fractional Replication

In the designs considered up to this point, all the possible treatment combinations in the factorial set were included in the actual experiment. The number of treatment combinations in a complete factorial set becomes quite large as the number of factors increases. For example, a 2^8 factorial experiment requires 256 treatment combinations; a 2^{16} factorial experiment requires 65,536 treatment combinations. If higher-order interactions can be considered negligible relative to main effects and lower-order interactions, only a selected fraction of the complete factorial set needs to be included in an experiment. The cost to the experimenter is the confounding of higher-order interactions with main effects and lower-order interactions. The gain is usually a substantial reduction in experimental effort, accompanied by a somewhat broader scope for the inferences. In cases where ambiguity is present because of confounding, the initial experiment may be supplemented by follow-up experiments specifically designed to clarify such ambiguities. There are situations in which it is highly

desirable to run a sequence of fractional replications, the choice of the successive fractions being determined by the results of the preceding fractions. In the designs to be discussed in this section, all factors are considered to be fixed.

The principles for selecting the set of treatments which will provide maximum information on main effects and lower-order interactions are essentially those followed in assigning treatments to blocks in a complete factorial experiment. To illustrate the kind of confounding which arises in a fractional replication, consider a one-half replication of a 2^3 factorial experiment. Suppose that the treatments in this one-half replication correspond to the treatments in the set $(ABC)_0$. The latter are (000), (011), (101), (110). Comparisons corresponding to the main effects and interactions may be indicated schematically as follows (the columns indicate the weights for a comparison):

	<i>A</i>	<i>B</i>	<i>C</i>	<i>AB</i>	<i>AC</i>	<i>BC</i>	<i>ABC</i>
(000) = (1)	—	—	—	—	—	—	—
(011) = <i>bc</i>	—	+	+	+	+	—	—
(101) = <i>ac</i>	+	—	+	+	—	+	—
(110) = <i>ab</i>	+	+	—	—	+	+	—

Note that the pattern of the comparison corresponding to the main effect of *A* (the column headed *A*) and the pattern of the comparison corresponding to the *BC* interaction (the column headed *BC*) are identical. (These two patterns would not continue to be identical if the remaining treatments in the complete factorial were added.) Hence, if only information on these four treatment combinations is available, variation due to the main effect of *A* is completely confounded with variation due to the *BC* interaction. Similarly, by using information from the one-half replication given above, *B* and *AC* are completely confounded; *C* and *AB* are also completely confounded. The sign pattern of *ABC* is not that of a comparison, it will be found that variation due to *ABC* cannot be estimated.

The effects which cannot be distinguished in a fractional replication are called aliases—the same source of variation is, in essence, called by two different names. The aliases in a fractional replication may be determined by means of a relatively simple rule. Consider a one-half replication of a 2^4 factorial experiment defined by the relation $(ABCD)_0$. The alias of the main effect of *A* is given by

$$A \times ABCD = A^2BCD = BCD.$$

BCD is the generalized interaction of *A* and *ABCD*; it is also the alias of *A* in a one-half replication defined by $(ABCD)_0$ or $(ABCD)_1$. The alias of *AB* is

$$AB \times ABCD = A^2B^2CD = CD.$$

The alias of ABC is the generalized interaction of ABC and the defining relation, i.e.,

$$ABC \times ABCD = D.$$

In general, the alias of an effect in a fractional replication is the generalized interaction of that effect with the defining relation or relations, should there be more than one.

The analysis of variance for a one-half replication of a 2^4 factorial experiment, by using $(ABCD)_0$ or $(ABCD)_1$ as the defining relation, is given in Table 8.15-1. The aliases are indicated. (It will be noted that, if BCD is the generalized interaction of A and $ABCD$, then A will be the generalized interaction of BCD and $ABCD$.) Main effects are aliased with three-

Table 8.15-1 One-half Replication of a 2^4 Factorial Experiment

Source	df
A (BCD)	1
B (ACD)	1
C (ABD)	1
D (ABC)	1
AB (CD)	1
AC (BD)	1
AD (BC)	1
Within cell	$8(n - 1)$
Total	$8n - 1$

factor interactions, and two-factor interactions are aliased with other two-factor interactions. The four-factor interaction cannot be estimated. If three-factor and higher-order interactions may be considered negligible, then estimates of the variance due to the main effects are given by this fractional replication. There is considerable ambiguity about what interpretation should be made if two-factor interactions should prove to be significant, since pairs of two-factor interactions are completely confounded.

By way of contrast, consider a one-half replication of a 2^5 factorial experiment. Suppose that the treatments are selected through use of the relation $(ABCDE)_0$ or $(ABCDE)_1$. The analysis of variance may be outlined as indicated below (aliases are indicated in parentheses):

Source	df
Main effects (4-factor interactions)	5
Two-factor interactions (3-factor interactions)	10
Within cell	$16(n - 1)$
Total	$16n - 1$

If three-factor and higher-order interactions are negligible, this fractional replication provides information for tests on main effects as well as two-factor interactions.

A single defining relation for a 2^k factorial experiment will select a one-half replication. If k is large, it is desirable to have a one-quarter or even a one-eighth replication. In order to select a one-fourth replication, the treatments are required to satisfy, simultaneously, two defining relations. Suppose that these defining relations are designated U and V . The aliases of an effect E will have the form

$$E \times U, \quad E \times V, \quad E \times (U \times V).$$

In words, the aliases of an effect in a one-fourth replication are the generalized interactions of that effect with each of the defining relations as well as the generalized interaction of the effect with the interaction of the defining relations.

For example, consider a one-fourth replication of a 2^6 factorial experiment. Suppose that the defining relations are $(ABCE)_0$ and $(ABDF)_0$. The generalized interaction of these two relations is $(CDEF)_0$. A partial list of the aliases is as follows:

Effect	Aliases
<i>A</i>	<i>BCE, BDF, ACDEF</i>
<i>B</i>	<i>ACE, ADF, BCDEF</i>
<i>AB</i>	<i>CE, DF, ABCDEF</i>
<i>CD</i>	<i>ABED, ABCF, EF</i>

In this case, main effects are aliased with three-factor and higher-order interactions; two factor interactions are aliased with other two-factor as well as higher-order interactions.

As another example, consider a one-fourth replication of a 2^7 factorial experiment. Suppose that the defining relations for the selected treatments are $(ABCDE)_0$ and $(ABCFG)_0$. The generalized interaction of these relations is $(DEFG)_0$. A partial listing of the aliases is as follows:

Effect	Aliases
<i>A</i>	<i>BCDE, BCFG, ADEFG</i>
<i>B</i>	<i>ACDE, ACFG, BDEFG</i>
<i>AB</i>	<i>CDE, CFG, ABDEFG</i>
<i>DE</i>	<i>ABC, ABCDEFG, FG</i>

In this design main effects are aliased with four-factor and higher-order interactions. With the exception of $DE = FG$, $DG = EF$, $DF = EG$, all other two-factor interactions are aliased with three-factor or higher-order interactions. The two-factor interactions which are equated are aliased.

This design yields unbiased tests on main effects and several of the two-factor interactions, provided that three-factor and higher-order interactions are negligible.

For a one-eighth replication of a 2^k factorial experiment, three defining relations are required to select the set of treatments. If these relations are designated U , V , and W , the aliases of an effect E are given by the following interactions:

$$E \times U, \quad E \times V, \quad E \times W;$$

$$E \times (U \times V), \quad E \times (U \times W), \quad E \times (V \times W), \quad E \times (U \times V \times W).$$

In general, if m is the number of defining relations, then the number of aliases of an effect is

$$m + \binom{m}{2} + \binom{m}{3} + \cdots + \binom{m}{m}.$$

For example, when $m = 3$,

$$3 + \binom{3}{2} + \binom{3}{3} = 3 + 3 + 1 = 7.$$

Fractional Replication in Blocks. To illustrate the general method for arranging the treatments in a fractional replication into blocks, consider

Table 8.15-2 One-half Replication of 2^6 Factorial
[Defining Relation: $(ABCDE F)_0$]

Block 1 (ABC) ₀				Block 2 (ABC) ₁			
(1)	ab	ac	bc	ae	af	ad	bd
abef	ef	de	df	bf	be	ce	cf
acde	acdf	abdf	acef	cd	abcd	abcf	abce
bcd f	bcde	bcef	abde	abcdef	cdef	bdef	adef
Block 1' (ABD) ₀		Block 1'' (ABD) ₁		Block 2' (ABD) ₁		Block 2'' (ABD) ₀	
(1)	ab	ac	bc	ae	af	ad	bd
abef	ef	de	df	bf	be	ce	cf
acde	acdf	abdf	acef	cf	abcd	abcf	abce
bcd f	bcde	bcef	abde	abcdef	cdef	bdef	adef

a 2^6 factorial experiment. Suppose that a one-half replication of 64 treatments is selected by means of the relation $(ABCDE F)_0$. Now suppose that the resulting 32 treatments are subdivided into two sets of 16 treatments by means of the relation ABC , one set being $(ABC)_0$ and the other $(ABC)_1$. The resulting sets of 16 treatments are given in part i of Table 8.15-2. Suppose that the latter sets define the blocks. In the analysis of the resulting experiment, main effects will be aliased with five-factor interactions. Two-factor interactions will be aliased with four-factor interactions.

Three-factor interactions will be aliased in pairs as follows:

Effect	Alias
ABC	DEF
ABD	CEF
ABE	CDF

The pair $ABC = DEF$ is confounded with differences between blocks.

Suppose that the blocks of 16 are further subdivided into blocks of size 8 by means of the relations $(ABD)_0$ and $(ABD)_1$. The resulting blocks are given in part ii of the table. The aliases remain the same as in the previous analysis, but now there is additional confounding with blocks. The three degrees of freedom confounded with between-block differences are

$$ABC = DEF, \quad ABD = CEF,$$

as well as

$$ABC \times ABD = CD.$$

Note that the generalized interaction of DEF and CEF is also CD .

Thus in a one-half replication of a 2^6 factorial experiment in blocks of size 8, if three-factor and higher-order interactions are negligible, main effects and all two-factor interactions except CD may be tested. Some of the three-factor interactions as well as CD are confounded with block effects, but other pairs of three-factor interactions are clear of block effects, provided that the appropriate model is strictly additive with respect to block effects.

Computational Procedures. Computational procedures for a one-half replication of a 2^k factorial experiment are identical to those of a complete replication of a 2^{k-1} factorial experiment. For example, the eight treatments in a one-half replication of a 2^4 factorial experiment actually form a complete replication of a 2^3 factorial experiment if the levels of one of the factors are disregarded. If the one-half replication is defined by the relation $(ABCD)_0$, the treatments are

$$(1), ab, ac, ad, bc, bd, cd, abcd.$$

If the levels of factor D are disregarded, the treatments are

$$(1), ab, ac, a, bc, b, c, abc;$$

these are the eight treatment combinations in a 2^3 factorial experiment. Corresponding effects in the three-factor and four-factor experiments are given below:

Three-factor experiment	Four-factor experiment (one-half rep)
A	$A = BCD$
B	$B = ACD$
C	$C = ABD$
AB	$AB = CD$
AC	$AC = BD$
BC	$BC = AD$
ABC	$ABC = D$

Suppose that a one-quarter replication of a 2^5 factorial experiment is defined by the relations $(ABE)_0$ and $(CDE)_0$. The treatments in this fractional replication are

$$(1), ab, cd, ace, bce, ade, bde, abcde.$$

If factors D and E are disregarded, the treatments are

$$(1), ab, c, ac, bc, a, b, abc;$$

these are the treatments in a complete replication of a 2^3 factorial experiment. The aliases of these effects are their generalized interactions with ABE , CDE , and $ABE \times CDE = ABCD$. Thus there is the following correspondence between this 2^3 factorial experiment and the one-quarter replication of a 2^5 factorial experiment:

Five-factor experiment (one-quarter rep)
$A = BE, ACDE, BCD$
$B = AE, BCDE, ACD$
$C = ABCE, DE, ABD$
$AB = E, ABCDE, CD$
$AC = BCE, ADE, BD$
$BC = ACE, BDE, AD$
$ABC = CE, ABDE, D$

That is, if a one-quarter replication of a 2^5 factorial is analyzed as if it were a complete replication of a 2^3 factorial, the corresponding effects are indicated—thus the analysis reduces to the analysis of a complete factorial having two fewer factors.

Extensive tables of fractional replications of experiments in the 2^k series are given in Cochran and Cox (1957, pp. 276–289). The plans in these tables permit the arrangement of the treatments into blocks of various sizes. To avoid having main effects and two-factor interactions aliased with lower-order interactions, care must be taken in selecting the defining relations for fractional replications and blocks. The plans tabulated in Cochran and Cox are those which tend to minimize undesirable aliases.

Fractional Replication for Designs in the 3^k Series. Principles used in the 2^k series may be generalized to factorial experiments in the p^k series, where p is a prime number. In a 3^3 factorial experiment, a one-third replication may be constructed by any of the components of the three-factor interaction: ABC , ABC^2 , AB^2C , or AB^2C^2 . Each of these components will subdivide the 27 treatments in a 3^3 factorial set into three sets of 9 treatments each. If one of the sets is defined by $(ABC)_i$, where $i = 0, 1$, or 2 is used, the aliases of the main effect of factor A are

$$A \times ABC = A^2BC = AB^2C^2,$$

$$A^2 \times ABC = BC.$$

The aliases of AB^2 are

$$AB^2 \times ABC = A^2B^3C = AC^2,$$

$$(AB^2)^2 \times ABC = A^3B^5C = B^2C = BC^2.$$

In general, if the defining relation for a one-third replication in a 3^k experiment is R , then the aliases of an effect E are

$$E \times R \quad \text{and} \quad E^2 \times R.$$

The following correspondence may be established between a one-third replication of a 3^3 factorial and a complete replication of a 3^2 factorial; assume that the defining relation is $(ABC)_i$:

3^2 factorial experiment	3^3 factorial experiment (one-third rep)
A	$A = AB^2C^2 = BC$
B	$B = AB^2C = AC$
AB	$AB = ABC^2 = C$
AB^2	$AB^2 = AC^2 = BC^2$

From the point of view of computational procedures, a one-third replication of a 3^3 factorial is equivalent to a complete replication of a 3^2 factorial experiment. Since two-factor interactions are aliased with main effects, this plan

Table 8.15-3 One-third Replication of a 3^4 Factorial Experiment
(Defining Relation: $ABCD$)

Source	df
$A = AB^2CD^2 = BCD$	2
$B = AB^2CD = ACD$	2
$C = ABC^2D = ABD$	2
$AB = ABC^2D^2 = CD$	2
$AB^2 = AC^2D^2 = BC^2D^2$	2
$AC = AB^2CD^2 = BD$	2
$AC^2 = AB^2D^2 = BC^2D$	2
$BC = AB^2C^2D = AD$	2
$BC^2 = AB^2D = AC^2D$	2
$ABC = ABCD^2 = D$	2
$ABC^2 = ABD^2 = CD^2$	2
$AB^2C = ACD^2 = BD^2$	2
$AB^2C^2 = AD^2 = BCD^2$	2
Within cell	27(n - 1)
Total	27n - 1

is of little practical use unless it can be assumed that all interactions are negligible.

A one-third replication of a 3^4 factorial experiment may be constructed from any one of the components of the four-factor interaction. If one of

the sets of 27 treatments defined by $(ABCD)_i$ is used, then the correspondence given in Table 8.15-3 may be established between the fractional replication of a 3^4 factorial and a complete replication of a 3^3 factorial. Main effects are seen to be aliased with three-factor and higher-order interactions. Some of the two-factor interactions are also aliased with other two-factor interactions. If, however, the two-factor interactions with factor D are negligible, then the other two-factor interactions are clear of two-factor aliases.

Assignment of treatments to blocks follows the same general principles as those given in earlier sections of this chapter. Plans for fractional replications in the 3^k series are given in Cochran and Cox (1957, pp. 290-291). The following notation systems are equivalent:

$$\begin{aligned} AB(I) &= AB^2, & ABC(W) &= AB^2C^2, & ABC(Y) &= ABC^2, \\ AB(J) &= AB, & ABC(X) &= AB^2C, & ABC(Z) &= ABC. \end{aligned}$$

Other Fractional-replication Designs. Latin squares and Greco-Latin squares may be considered as fractional replications of factorial experiments. Experimental designs in these categories will be considered in Chap. 10.

CHAPTER 9

Balanced Lattice Designs and Other Balanced Incomplete-block Designs

9.1 General Purpose

The designs discussed in Chap. 8 permit the experimenter to assign the treatment combinations in a factorial experiment to blocks of size k , where k is smaller than the pq treatment combinations in a two-factor factorial experiment. In balanced designs, some within-block information is available on all main effects and interactions, the information on the interactions being in some cases only partially complete.

The design principles developed in Chap. 8 may be adapted for use in assigning the p treatments in a single-factor experiment to blocks of size k , where k is smaller than p . In particular, the principles developed in the last chapter are useful in constructing balanced lattice designs. In an m -dimensional balanced lattice the number of treatments is k^m , where k is a prime number. For example, in a balanced two-dimensional (double) design there are k^2 treatments. In a balanced three-dimensional (cubic) lattice there are k^3 treatments. The block size in a balanced lattice of dimension m is k or k to some power less than m .

Lattice designs are also classified according to the number of dimensions which are in balance, i.e., the number of restrictions on the structure of the design. For example, a balanced double lattice may be balanced in only one dimension (one restriction) or balanced in both dimensions (two restrictions). A cubic lattice may have one, two, or three restrictions. In this chapter balanced double lattice designs having one and two restrictions will be considered. For partially balanced lattice designs as well as higher-order balanced lattice designs the reader is referred to an extensive treatment of these topics given in Federer (1955).

Balanced lattice designs are special cases of what are known as balanced incomplete-block designs. In the latter designs, t treatments are arranged in blocks of size k in such a way that unbiased estimates of treatment effects may be obtained. These designs are particularly useful in those situations

requiring within-subject estimates of treatment effects without the requirement that each subject be observed under all treatments. Each subject is required to be observed only under k of the p treatments. The analysis of balanced incomplete-block designs assumes that a strictly additive model is appropriate; i.e., no interaction terms are included in the model. Also, in these designs the block effects are assumed to be random, and the treatment effects are assumed to be fixed. However, all the assumptions about constant correlations in a repeated-measures experiment must be met if repeated measures are used.

9.2 Balanced Simple Lattice

A balanced double lattice having one restriction is called a simple lattice. In this design k^2 treatments are arranged in blocks of size k , where k is a prime number. The treatments must be assigned to the blocks in such a way that each possible pair of treatments occurs together in some block an equal number of times. If this condition is met, the within-block information on all pairs of treatments is equal—in this sense, the design is balanced. The minimum number of blocks required to achieve balance is $k(k + 1)$.

The construction of the blocks is simplified by associating each of the k^2 treatments with one of the treatment combinations in a $k \times k$ factorial experiment. The blocks are then constructed in terms of the treatment combinations of the factorial experiment. In terms of this latter type of experiment, the effects confounded with blocks are more readily identified.

A simple balanced lattice in which $k^2 = 9$ will be used for illustrative purposes. Consider the design given in part i of Table 9.2-1. Differences between blocks within replication 1 are confounded with the main effect of pseudo factor A . Blocks in replication 2 are confounded with pseudo factor B . In replications 3 and 4, differences between blocks are confounded, respectively, with the AB and AB^2 components of the interaction. Thus, in terms of the pseudo factors, there is $\frac{3}{4}$ within-block information on each of the main effects as well as $\frac{3}{4}$ within-block information on each of the components of the interaction. In general the within-block information per component of a pseudo factor is $k/(k + 1)$.

The translation from a design in terms of pseudo factors to a design in terms of treatments in a single-factor experiment is given in part ii. Any mode of associating the symbol (ij) with a treatment may be used so long as the method of association is consistent for all replications. In part ii it may be verified that each possible pair of treatments occurs in the same block once and only once. For example, the pair (1,5) occurs only in block 10; the pair (7,9) occurs only in block 3. Balanced lattices for $k^2 = 9, 16, 25, 49, 64$, and 81 have been constructed. No balanced lattice exists for $k^2 = 36$; none has been constructed for $k^2 = 100$ or 144.

Since all the treatments do not appear in the same block, treatment effects are partially confounded with block effects. Conversely, block effects are

Table 9.2-1 Balanced Simple Lattice for Nine Treatments in Blocks of Size 3

of Size 3

Block	Replication 1				Block	Replication 2			
1	(00)	(01)	(02)	$x_1 = 0$	4	(00)	(10)	(20)	$x_2 = 0$
2	(10)	(11)	(12)	$x_1 = 1$	5	(01)	(11)	(21)	$x_2 = 1$
3	(20)	(21)	(22)	$x_1 = 2$	6	(02)	(12)	(22)	$x_2 = 2$
(i) Block	Replication 3				Block	Replication 4			
7	(00)	(12)	(21)	$x_1 + x_2 = 0$	10	(00)	(11)	(22)	$x_1 + 2x_2 = 0$
8	(01)	(10)	(22)	$x_1 + x_2 = 1$	11	(02)	(10)	(21)	$x_1 + 2x_2 = 1$
9	(02)	(11)	(20)	$x_1 + x_2 = 2$	12	(01)	(12)	(20)	$x_1 + 2x_2 = 2$

Block	Replication 1			Block	Replication 2		
1	1	2	3	4	1	4	7
2	4	5	6	5	2	5	8
3	7	8	9	6	3	6	9
(ii) Block	Replication 3			Block	Replication 4		
7	1	6	8	10	1	5	9
8	2	4	9	11	3	4	8
9	3	5	7	12	2	6	7

also partially confounded with treatment effects. The primary purpose of this type of design is to obtain estimates of treatment effects which are not biased by block effects. To achieve this objective, one proceeds in a rather roundabout manner—estimates of block effects which are not biased by treatment effects are obtained first. This is done in a preliminary analysis of variance which takes the following general form:

Source	df	MS
Replications	k	
Treatments (unadj)	$k^2 - 1$	
Blocks (adj) w. rep	$(k - 1)(k + 1)$	E_b
Intrablock error	$(k - 1)(k^2 - 1)$	E_e
Total	$k^3 + k^2 - 1$	

(It is to be noted that the blocks are nested under the replications.) In the analysis given above, the mean square due to blocks adjusted for treatment effects is generally designated by the symbol E_b ; the latter mean square is also referred to as the between-block error term. Similarly, the symbol E_e is widely used to indicate the within-block, or intrablock, error term. The expected value of these mean squares has the following general form:

$$E(E_b) = \sigma_e^2 + \frac{k^2}{k + 1} \sigma_{\text{blocks}}^2; \quad E(E_e) = \sigma_e^2.$$

Some of the principles underlying the adjustment of the block effects will be illustrated in terms of the design given in Table 9.2-1. Toward this end, the following notation will be used:

$$\begin{aligned} T_i &= \text{sum of the } k + 1 = 4 \text{ observations on treatment } i. \\ B_j &= \text{sum of the } k = 3 \text{ observations within block } j. \\ B_{(i)} &= \sum_{(i)} B_j = \text{sum of the block totals in which treatment } i \text{ appears.} \end{aligned}$$

In terms of a strictly additive model,

$$T_1 = 4\mu + 4\tau_1 + \beta_1 + \beta_4 + \beta_7 + \beta_{10}.$$

(Terms involving ε are not included.) Since treatment 1 appears in blocks 1, 4, 7, and 10, the corresponding block effects appear on the right-hand side of the above expression in addition to the direct effect associated with treatment 1. The totals for the blocks in which treatment 1 appears estimate the following parameters of a strictly additive model:

$$\begin{aligned} B_1 &\doteq 3\mu + \tau_1 + \tau_2 + \tau_3 + 3\beta_1, \\ B_4 &\doteq 3\mu + \tau_1 + \tau_4 + \tau_7 + 3\beta_4, \\ B_7 &\doteq 3\mu + \tau_1 + \tau_6 + \tau_8 + 3\beta_7, \\ B_{10} &\doteq 3\mu + \tau_1 + \tau_5 + \tau_9 + 3\beta_{10}. \end{aligned}$$

In the model, $\sum \tau_i = 0$; hence,

$$B_{(1)} = B_1 + B_4 + B_7 + B_{10} \doteq 12\mu + 3\tau_1 + 3(\beta_1 + \beta_4 + \beta_7 + \beta_{10}).$$

The grand total of the $k^2(k + 1) = 36$ observations in the experiment provides an estimate of 36μ . Hence,

$$W_1 = 3T_1 - 4B_{(1)} + G \doteq -9(\beta_1 + \beta_4 + \beta_7 + \beta_{10}).$$

That is, W_1 is a function of the sum of block effects in which treatment 1 appears. In general,

$$W_i = kT_i - (k + 1)B_{(i)} + G \doteq -k^2 \sum_{(i)} \beta_j.$$

Thus each W_i provides an estimate of a quantity which is a function only of block effects.

An estimate of the effect of treatment i , adjusted for block effects, is given by

$$Q_i = kT_i - B_{(i)} \doteq k^2\tau_i.$$

If there were no confounding,

$$kT_i - k(k + 1)\bar{G} \doteq k(k + 1)\tau_i.$$

The relative within-block information, or efficiency, is

$$\text{Eff} = \frac{k^2}{k(k + 1)} = \frac{k}{k + 1}.$$

The quantity Q_i provides an intrablock estimate of treatment effect τ_i .

The estimate of variation due to block effects, adjusted for treatments, is given by

$$SS_{\text{blocks(adj)}} = \frac{\sum W_i^2}{k^3(k+1)}.$$

The expected value of the mean square corresponding to this source of variation may be shown to be

$$E(E_b) = \sigma_e^2 + \frac{k^2}{k+1} \sigma_\beta^2.$$

The other sources of variation in the preliminary analysis are obtained as follows: If R_m is the sum of the k^2 observations in replication m ,

$$SS_{\text{reps}} = \frac{\sum R_m^2}{k^2} - \frac{G^2}{k^2(k+1)}.$$

The variation due to the treatments, the blocks being disregarded, is

$$SS_{\text{treat(unadj)}} = \frac{\sum T_i^2}{k+1} - \frac{G^2}{k^2(k+1)}.$$

The total variation is given by

$$SS_{\text{total}} = \sum X^2 - \frac{G^2}{k^2(k+1)}.$$

The intrablock error is obtained by subtraction.

$$SS_{\text{inblock}} = SS_{\text{total}} - SS_{\text{reps}} - SS_{\text{treat(unadj)}} - SS_{\text{blocks(adj)}}.$$

Information for estimating treatment effects comes from two sources. The primary source is that obtained from differences between pairs of treatments in the same block—this source is known as the intrablock, or within-block, information. The error associated with estimates obtained from intrablock information is E_e , the intrablock mean square. The secondary source of information on the treatment effects is obtained from differences between the $B_{(i)}$'s; such information is called interblock treatment information. The error associated with this source of information is E_b . (Usually E_b will be considerably larger than E_e .)

In general, given two estimates m_1 and m_2 of the same parameter with respective variances σ_1^2 and σ_2^2 , the best linear unbiased estimate of the parameter has the form

$$m = \frac{(m_1/\sigma_1^2) + (m_2/\sigma_2^2)}{(1/\sigma_1^2) + (1/\sigma_2^2)}.$$

In words, the best unbiased linear estimate is a weighted average of the individual estimates, the weights being the reciprocals of the respective variances. This last expression is algebraically equivalent to

$$m = m_1 + \mu(m_2 - m_1),$$

where

$$\mu = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}.$$

If σ_2^2 is large relative to σ_1^2 , μ will tend to be near zero. In this case m_1 will not differ appreciably from m . The conditions under which m is a uniformly better unbiased estimator of the parameter estimated by m_1 and m_2 are discussed by Graybill and Deal (1959).

This general principle for combining two estimates may be used in combining intrablock and interblock treatment information. In the latter case the best linear unbiased estimate of the sum of the effects due to treatment i has the form

$$T'_i = T_i + \mu W_i, \quad \text{where } \mu = \frac{E_b - E_e}{k^2 E_e}.$$

In this last expression, the term W_i is not of the form $m_2 - m_1$; hence μ does not have the same definition as that given in the last paragraph. (Since the parameter which μ estimates is a ratio of variances, the parameter cannot assume a negative value. If μ should be negative, μ is assumed to be zero.) When E_b is large relative to E_e , only the within-block information on the treatments need be taken into account. As estimate of the latter is given by

$$T''_i = \frac{k+1}{k^2} Q_i + (k+1)\bar{G}.$$

The error variance associated with this estimate is E_e .

The error variance associated with estimates which utilize both intra- and interblock information is

$$E'_e = E_e(1 + k\mu).$$

E'_e is the effective error mean square, and its degrees of freedom are taken as being numerically equal to the degrees of freedom of E_e . Individual tests between adjusted treatment totals have the form

$$F = \frac{(T'_1 - T'_2)^2}{2(k+1)E'_e}.$$

Any of the procedures for making all possible comparisons among treatment means may be adapted for use here—in this case the adjusted treatment means are used, and E'_e has the role of the error mean square.

An approximate over-all test on the differences between treatments uses

$$SS_{\text{treat(adj)}} = \frac{\sum (T'_i)^2}{k+1} - \frac{G^2}{k^2(k+1)}.$$

The F ratio has the form

$$F = \frac{MS_{\text{treat(adj)}}}{E'_e}.$$

A simple balanced lattice for the case $k^2 = 16$ in blocks of size $k = 4$ is given in Table 9.2-2. The number of replications required for balance is $k+1 = 5$. Note that the pair of treatments (1,2) occurs together in the

Table 9.2-2 Balanced Simple Lattice for 16 Treatments in Blocks of Size 4

Block	Replication 1				Block	Replication 2			
1	1	2	3	4	6	1	5	9	13
2	5	6	7	8	7	2	6	10	14
3	9	10	11	12	8	3	7	11	15
4	13	14	15	16	9	4	8	12	16
Block	Replication 3				Block	Replication 4			
9	1	6	11	16	13	1	14	7	12
10	5	2	15	12	14	13	2	11	8
11	9	14	3	8	15	5	10	3	16
12	13	10	7	4	16	9	6	15	4
		Block	Replication 5						
		17	1	10	15	8			
		18	9	2	7	16			
		19	13	6	3	12			
		20	5	14	11	4			

same block only once—namely, in block 1. Note also that the pair (6,16) occurs together in the same block only once—namely, in block 9. In general, each of the $[16(15)]/[1(2)] = 120$ possible unique pairs occurs together in the same block once and only once. Since each block contains $[4(3)]/[1(2)] = 6$ pairs, the minimum number of blocks required for balance is $120/6 = 20$.

In this design there are $k + 1 = 5$ observations on each treatment. If it is desired to have $5n$ observations on each treatment, n sets of $k + 1$ replications may be constructed. The latter design would require $20n$ subjects. More extensive tables of balanced simple lattices are given in Cochran and Cox (1957, pp. 428–432). In using these designs, the blocks within a replication are arranged in a randomized order, and the treatments within a block are also arranged in a randomized order.

The blocks in a balanced simple lattice can always be arranged in distinct replications. When, for example, a sample of 20 subjects is assigned at random to the blocks in the design given in Table 9.2-2, the replication dimension of the design does not exist. If the replications are disregarded, the analysis of variance has the following form:

Source	df	MS
Treatments (unadj)	$k^2 - 1$	E_b E_e
Blocks (adj)	$k(k + 1) - 1$	
Intrablock error	$(k - 1)(k^2 - 1)$	
Total	$k^2 + k^2 - 1$	

When the replications are disregarded, the adjusted sum of squares for blocks is given by

$$SS_{\text{blocks(adj)}} = SS_{\text{blocks(unadj)}} + SS_{\text{treat(adj for blocks)}} - SS_{\text{treat(unadj)}}$$

The unadjusted sums of squares for blocks and treatments are computed as if there were no confounding; i.e.,

$$SS_{\text{blocks(unadj)}} = \frac{\sum B_j^2}{k} - \frac{G^2}{k^2(k+1)},$$

$$SS_{\text{treat(unadj)}} = \frac{\sum T_i^2}{k+1} - \frac{G^2}{k^2(k+1)}.$$

To obtain the sum of squares for treatments adjusted for blocks, one first computes quantities of the form

$$Q_i = kT_i - B_{(i)}.$$

Then
$$SS_{\text{treat(adj for blocks)}} = \frac{\sum Q_i^2}{k^3}.$$

The adjusted treatment totals, utilizing both intrablock and interblock information, are given by

$$T'_i = T_i + \mu' W_i,$$

where
$$\mu' = \frac{(k^2 + k - 1)(E_b - E_r)}{k^2[(k^2 + k - 1)E_b + E_r]}.$$

It should be noted that $SS_{\text{treat(adj for blocks)}}$ is different from $SS_{\text{treat(adj)}}$. The latter utilizes both intrablock and interblock information; the former utilizes only intrablock information. In this case

$$SS_{\text{treat(adj)}} = \frac{\sum (T'_i)^2}{k+1} - \frac{G^2}{k^2(k+1)}.$$

The effective error in this case is

$$E'_e = E_e[1 + k\mu'].$$

9.3 Numerical Example of Balanced Simple Lattice

Computational procedures for this design will be illustrated by means of an experiment having $k^2 = 9$ treatments. Suppose that the "treatments" are nine different pictures in a projective test. It is not considered experimentally feasible to have each subject respond to all the pictures. Each subject can, however, respond to three pictures. The criterion is the number of responses in a specified category. A balanced simple lattice design in which the subject corresponds to the block was chosen for use. The design

is indicated in Table 9.3-1. The numbers in parentheses indicate the treatments.

Four random samples of three subjects each were obtained for the experiment. The samples were assigned at random to the replications; subjects within samples were assigned at random to the blocks. Data obtained are shown in Table 9.3-1. For example, subject 1 responded to pictures (1), (3), and (2) and had respective criterion scores of 60, 95, and 90. The

Table 9.3-1 Numerical Example of Balanced Simple Lattice
($k = 3$)

Subject					Subject				
1	(1)	(3)	(2)	245	4	(4)	(1)	(7)	365
	60	95	90			160	115	90	
2	(6)	(4)	(5)	415	5	(2)	(5)	(8)	365
	150	125	140			120	160	85	
3	(8)	(7)	(9)	205	6	(9)	(6)	(3)	385
	70	75	60			100	175	110	
Rep 1 total =				865	Rep 2 total =				1115

Subject					Subject				
7	(5)	(1)	(9)	275	10	(1)	(8)	(6)	330
	140	85	50			105	75	150	
8	(2)	(7)	(6)	245	11	(2)	(4)	(9)	310
	70	65	110			100	130	80	
9	(8)	(3)	(4)	175	12	(5)	(7)	(3)	375
	20	60	95			170	85	120	
Rep 3 total =				695	Rep 4 total =				1015

$$\Sigma X^2 = 426,000$$
$$G^2/36 = 378,225$$
$$SS_{\text{total}} = 47,775$$

sequence in which the pictures were presented was determined by a randomization procedure. Block totals as well as replication totals are given in Table 9.3-1.

Totals needed to obtain sums of squares are given in Table 9.3-2. The first entry in column T_i is the sum of all observations on treatment 1; this total is

$$T_1 = 365$$

60 from block 1
 115 from block 4
 85 from block 7
 105 from block 10

Other entries in column T_i are obtained in an analogous manner. The first entry in column $B_{(i)}$ is

$$\begin{aligned} B_{(1)} &= \text{sum of totals for blocks containing treat 1} \\ &= B_1 + B_4 + B_7 + B_{10} = \sum_{(1)} B_j \\ &= 245 + 365 + 275 + 330 = 1215. \end{aligned}$$

The second entry in column $B_{(i)}$ is

$$\begin{aligned} B_{(2)} &= \text{sum of totals for blocks containing treat 2} \\ &= B_1 + B_5 + B_8 + B_{11} = \sum_{(2)} B_j \\ &= 245 + 365 + 245 + 310 = 1165. \end{aligned}$$

The symbol $\sum_{(2)} B_j$ indicates the sum of block totals which contain an entry from treatment 2. As a partial check on the computations, the sum of the entries in column $B_{(i)}$ should be $(k-1)G$.

Table 9.3-2 Numerical Example (Continued)

Treat	T_i	$B_{(i)}$	$W_i = 3T_i - 4B_{(i)} + G$	$T'_i = T_i + \mu W_i$	$Q_i = 3T_i - B_{(i)}$	T''_i
1	365	1,215	-75	361.0	-120	356.7
2	380	1,165	170	389.0	-25	398.9
3	385	1,180	125	391.6	-25	398.9
4	510	1,265	160	518.5	265	527.8
5	610	1,430	-200	599.4	400	587.8
6	585	1,375	-55	582.1	380	578.9
7	315	1,190	-125	308.4	-245	301.1
8	250	1,075	140	257.4	-325	265.6
9	290	1,175	-140	282.6	-305	274.4
	$G = 3690$	$11,070 = 3G$	0	3690.0	0	3690.1

The first entry in column W_i is

$$\begin{aligned} W_1 &= kT_1 - (k+1)B_{(1)} + G \\ &= 3(365) - 4(1215) + 3690 = -75. \end{aligned}$$

Other entries in column W_i are obtained in an analogous manner. As a partial check on the computations the sum of the entries in column W_i should be zero.

The variation due to replications is

$$\begin{aligned} \text{SS}_{\text{reps}} &= \frac{\sum R_m^2}{k^2} - \frac{G^2}{k^2(k+1)} \\ &= \frac{856^2 + 1115^2 + 695^2 + 1015^2}{9} - \frac{3690^2}{36} \\ &= 11,186. \end{aligned}$$

The variation due to treatments, the confounding with blocks being disregarded, is

$$\begin{aligned} SS_{\text{treat(unadj)}} &= \frac{\sum T_i^2}{k+1} - \frac{G^2}{k^2(k+1)} \\ &= \frac{365^2 + 380^2 + \cdots + 290^2}{4} - \frac{3690^2}{36} \\ &= 33,300. \end{aligned}$$

The variation due to blocks, adjusted for confounding with treatments, is

$$\begin{aligned} SS_{\text{blocks(adj)}} &= \frac{\sum W_i^2}{k^3(k+1)} - \frac{(-75)^2 + (170)^2 + \cdots + (-140)^2}{3^3(4)} \\ &= 1607. \end{aligned}$$

The total variation is

$$SS_{\text{total}} = \sum X^2 - \frac{G^2}{k^2(k+1)} = 426,000 - 378,225 = 47,775.$$

The variation due to intrablock error is obtained by subtraction; i.e.,

$$\begin{aligned} SS_{\text{intrablock error}} &= SS_{\text{total}} - SS_{\text{reps}} - SS_{\text{treat(unadj)}} - SS_{\text{blocks(adj)}} \\ &= 1682. \end{aligned}$$

A summary of the analysis of variance is given in Table 9.3-3. Since the blocks are nested within the replications the degrees of freedom are $(k-1)(k+1) = 8$.

There is an alternative, more general, computational procedure for obtaining the sum of squares for blocks adjusted for treatments. It will be found that

$$SS_{\text{blocks(adj)}} = SS_{\text{blocks w. reps}} + SS_{\text{treat(adj for blocks)}} - SS_{\text{treat(unadj)}}.$$

Variation due to blocks within replications, confounding with treatments being disregarded, is

$$\begin{aligned} SS_{\text{blocks w. reps}} &= \frac{\sum B_j^2}{k} - \frac{\sum R_m^2}{k+1} \\ &= 10,872. \end{aligned}$$

Variation due to treatments adjusted for blocks is given by

$$SS_{\text{treat(adj for blocks)}} = \frac{\sum Q_i^2}{k^3} = 24,035.$$

Thus,

$$SS_{\text{blocks(adj)}} = 10,872 + 24,035 - 33,300 = 1607.$$

Within rounding error, the two computational procedures will give identical results.

The weighting factor μ needed to obtain an estimate of the adjusted treatment total T'_i is

$$\mu = \frac{E_b - E_e}{k^2 E_b} = \frac{201 - 105}{9(201)} = .0531.$$

The adjusted treatment totals are given in Table 9.3-2. For example,

$$T'_1 = T_1 + \mu W_1 = 365 + (.0531)(-75) = 361.0.$$

This total combines both intrablock and interblock treatment information. An estimate which uses only intrablock information is given by

$$\begin{aligned} T''_1 &= \frac{k+1}{k^2} Q_1 + (k+1)\bar{G} \\ &= \left(\frac{4}{9}\right)(-120) + 410 = 356.7. \end{aligned}$$

When E_b is large relative to E_e , T'_i and T''_i will be approximately equal.

Table 9.3-3 Summary of Analysis of Variance

Source of variation	SS	df	MS
Replications	11,186	3	4162
Treatments (unadj)	33,300	8	
Blocks within reps (adj)	1,607	8	
Intrablock error	1,682	16	
Total	47,775	35	

Tests on the difference between two T'_i 's, which are equivalent to tests on the difference between corresponding means, require the effective mean square for error,

$$E'_e = E_e(1 + k\mu) = 105[1 + 3(.0531)] = 122.$$

To test the hypothesis that $\tau_4 = \tau_8$ against a two-tailed alternative, one obtains the statistic

$$F = \frac{(T'_4 - T'_8)^2}{2(k+1)E'_e} = \frac{(518.5 - 257.4)^2}{8(122)} = 6.99.$$

For a .05-level test, the critical value is $F_{.95}(1,16) = 4.49$, assuming that this test is considered as being a priori. Since the observed F exceeds the critical value, the experimental data reject the hypothesis that $\tau_4 = \tau_8$.

An approximate over-all test on the hypothesis that $\sigma_\tau^2 = 0$ is given by

$$F = \frac{MS_{\text{treat(adj)}}}{E'_e}.$$

The sum of squares corresponding to the T'_i 's is

$$\begin{aligned} SS_{\text{treat(adj)}} &= \frac{\sum (T'_i)^2}{k+1} - \frac{G^2}{k^2(k+1)} \\ &= 32,571. \end{aligned}$$

From this,

$$MS_{\text{treat(adj)}} = \frac{SS_{\text{treat(adj)}}}{k^2 - 1} = 4071.$$

Hence

$$F = \frac{4071}{122} = 33.37.$$

The critical value for a .05-level test is $F_{.95}(8,16) = 2.59$. Any of the procedures for making all possible comparisons between the adjusted treatment means, $\bar{T}'_i = T'_i/(k+1)$, may be adapted for use here, with E'_e having the role of MS_{error} .

9.4 Balanced Lattice-square Designs

A balanced simple lattice has balance with respect to the rows (blocks) but not with respect to the columns (position within the block). A *balanced lattice square* has balance with respect to both rows and columns. The latter designs are readily constructed by using the same kind of pseudo factors introduced in connection with simple lattice designs. Principles underlying the analysis of lattice squares are also most readily understood in terms of these pseudo factors.

An example of a balanced lattice-square design is given in Table 9.4-1.

Table 9.4-1 Balanced Lattice Square

Replication 1			Replication 2		
(6)	(4)	(5)	(8)	(4)	(3)
(3)	(1)	(2)	(6)	(2)	(7)
(9)	(7)	(8)	(1)	(9)	(5)
Replication 3			Replication 4		
(4)	(2)	(9)	(7)	(1)	(4)
(3)	(7)	(5)	(9)	(3)	(6)
(8)	(6)	(1)	(8)	(2)	(5)

There are $k^2 = 9$ treatments arranged in 3×3 squares. A minimum of $k+1 = 4$ replications is required for double balance. To increase the number of observations on each treatment, a basic balanced plan may be repeated n times. In Table 9.4-1, each treatment is paired within the same row with every other treatment once. Further, each treatment is paired within the same column with every other treatment once. For example, the

following pairs, which include treatment 1, appear in the columns indicated:

(1,4) and (1,7)	Col. 2 of replication 1
(1,6) and (1,8)	Col. 1 of replication 2
(1,5) and (1,9)	Col. 3 of replication 3
(1,2) and (1,3)	Col. 2 of replication 4

Thus each of the other eight treatments is paired with treatment 1 in the same column. Balanced lattice-square designs may be constructed for k equal to a prime number or some power of a prime number; that is, $k = 2, 2^2, 3, 7, 9, 11, 13$. Plans will be found in Cochran and Cox (1957, pp. 497-506).

The analysis of variance for n repetitions of a basic balanced set is generally carried out in two steps. In the first step, the total variation is partitioned as follows:

Source	df
Replications	$n(k+1) - 1$
Treatments (unadj)	$k^2 - 1$
Residual	$(k^2 - 1)(kn + n - 1)$
Total	$nk^2(k+1) - 1$

In the second step, the residual variation is partitioned in two different ways. The first partition of the residual is as follows:

Source	df	MS
Residual	$(k^2 - 1)(kn + n - 1)$	
(a) Rows (mp) adj for treat	$n(k+1)(k-1)$	E_e
(b) Col. (mp) adj for treat and rows	$n(k+1)(k-1)$	
Intrablock error	$(kn - n - 1)(k^2 - 1)$	

The notation (mp) indicates that the rows and columns are considered to be nested within replication m in repetition p . Within a single basic balanced plan there are $k - 1$ degrees of freedom for the rows within each of the $k + 1$ replications—hence there are $(k + 1)(k - 1)$ degrees of freedom for rows. For n repetitions of the basic plan, there will be $n(k + 1)(k - 1)$ degrees of freedom for rows.

From this partition one obtains an estimate of variance due to the columns, E_c , which is free of treatment and row effects.

The second partition of the residual is made as follows:

Source	df	MS
Residual	$(k^2 - 1)(kn + n - 1)$	
(c) Cols. (mp) adj for treat	$n(k+1)(k-1)$	E_r
(d) Rows (mp) adj for treat and cols.	$n(k+1)(k-1)$	
Intrablock error	$(kn - n - 1)(k^2 - 1)$	

This partition provides an estimate of variance due to row effects which is free of treatment and column effects. Both partitions provide the same estimate of intrablock error. Since treatment effects are not orthogonal to rows and columns, the adjusted variations are not completely additive. In terms of the notation in the above partitions of the residual variation, there is, however, the following additivity:

$$(a) + (b) = (c) + (d).$$

An adjusted treatment total in a lattice-square design has the following form:

$$T'_i = T_i + \lambda' L_i + \mu' M_i.$$

where L'_i and M'_i depend upon row effects and column effects, respectively, and λ' and μ' are weighting factors which depend primarily upon E_r , E_c , and E_e . Adjusted totals of this kind utilize intra- and interrow and column treatment information. The expected values for the error terms are given in the discussion of the numerical example.

Numerical Example of Balanced Lattice Square. The example to be discussed will be cast in terms of a repeated-measure design. The rows of the square will represent the set of treatments assigned to a subject; the columns will represent the order in which the treatments are administered. Suppose that the basic data are those given in Table 9.4-2.

In this table there are $k^2 = 9$ treatments arranged in a balanced set of 3×3 squares. There are $n = 2$ repetitions of the same basic plan. Two samples of 12 subjects each are drawn at random from a specified population. The subjects from one sample are assigned at random to the rows of repetition a ; the subjects from the second sample are assigned at random to the rows of repetition b . The criterion scores for each of the subjects are indicated. For example, subject $1a$ has scores of 2, 3, and 7 under the respective treatments (6), (4), and (5).

The following notation will be used in the course of the analysis of variance:

- $T_{i(p)}$ = sum of $(k + 1)$ observations on treatment i in repetition p .
- $R_{j(m)p}$ = sum of k observations in row j within replication m of repetition p .
- $C_{j(m)p}$ = sum of k observations in column j within replication m of repetition p .
- $\text{Repl}_{m(p)}$ = sum of k^2 observations in replication m in repetition p .
- G_p = sum of $k^2(k + 1)$ observations in repetition p .
- $G = \Sigma G_p$ = sum of all observations in the experiment.
- $R_{i(p)}$ = sum of totals for all rows in repetition p which contain treatment i .
- $C_{i(p)}$ = sum of totals for all columns in repetition p containing treatment i .

The entry R_1 in repetition a is obtained as follows:

$$\begin{array}{r}
 14 \text{ row total in rep } 1a \text{ containing treat } 1 \\
 20 \text{ row total in rep } 2a \text{ containing treat } 1 \\
 11 \text{ row total in rep } 3a \text{ containing treat } 1 \\
 14 \text{ row total in rep } 4a \text{ containing treat } 1 \\
 \hline
 R_1 = 59
 \end{array}$$

The entry C_1 is obtained as follows:

$$\begin{array}{r}
 16 \text{ col total in rep } 1a \text{ containing treat } 1 \\
 13 \text{ col total in rep } 2a \text{ containing treat } 1 \\
 16 \text{ col total in rep } 3a \text{ containing treat } 1 \\
 13 \text{ col total in rep } 4a \text{ containing treat } 1 \\
 \hline
 C_1 = 58
 \end{array}$$

By using different sections of the keyboard on a calculator, T_1 , R_1 , and C_1 may be obtained in a single summation process. As a partial check on the computational work,

$$\Sigma T_{i(p)} = G_p; \quad \Sigma R_{i(p)} = \Sigma C_{i(p)} = kG_p.$$

It may be shown algebraically that

$$\begin{aligned}
 M'_{i(p)} &= D_{i(p)} + K_{i(p)} \\
 &= kT_{i(p)} - (k+1)C_{i(p)} + G_p.
 \end{aligned}$$

For example,
$$\begin{aligned}
 M'_{1(a)} &= 1 + 0 = 1 \\
 &= 3(29) - 4(58) + 146 = 1
 \end{aligned}$$

This dual method for computing M'_i also provides a numerical check.

The summary data in Table 9.4-4 may be computed by two different methods. The simplest method is to add corresponding entries in repetitions a and b in Table 9.4-3. Equivalently, one may compute entries in the first three columns directly from the basic data in Table 9.4-2 by summing over all repetitions. The other columns are obtained from the first three columns.

Convenient computational symbols are defined in Table 9.4-5. The basic data for the computation of symbols (5) through (8) are contained in Table 9.4-3. The summations are over all the repetitions. Data for symbols (2) and (4) are in Table 9.4-2; data for symbol (3) are in Table 9.4-4. The first phase of the analysis of variance appears at the bottom of Table 9.4-5. In reality the residual variation is obtained by subtracting the variation due to replications and treatment (unadjusted) from the total variation. This residual includes variation due to the columns, rows, and

Table 9.4-4 Summary Data for Numerical Example (Continued)

Treatment	Repetition a + repetition b							
	T_i	R_i	C_i	D_i	L'_i	J_i	K_i	M'_i
1	61	130	132	-2	9	7	3	1
2	27	106	112	-6	3	-3	-15	-21
3	38	110	110	0	20	20	20	20
4	32	113	120	-7	-10	-17	-31	-38
5	49	122	127	-5	5	0	-10	-15
6	27	102	104	-2	19	17	13	11
7	34	114	109	5	-8	-3	7	12
8	23	105	110	-5	-5	-10	-20	-25
9	55	136	114	22	-33	-11	33	55
	346 = G	1038	1038	0	0	0	0	0

Table 9.4-5 Computation of Sums of Squares

ANALYSIS OF VARIANCE

$$\begin{aligned}
 (1) &= G^2/k^2n(k+1) & (5) &= [\Sigma(L'_{i(p)})^2]/k^3(k+1) \\
 (2) &= \Sigma X^2 & (6) &= [\Sigma(K_{i(p)})^2]/k^3(k-1) \\
 (3) &= (\Sigma T_i^2)/n(k+1) & (7) &= [\Sigma(M'_{i(p)})^2]/k^3(k+1) \\
 (4) &= [\Sigma(\text{repl}_{m(p)})^2]/k^2 & (8) &= [\Sigma(J_{i(p)})^2]/k^3(k-1)
 \end{aligned}$$

$$\begin{aligned}
 (1) &= (346)^2/72 \\
 (2) &= \Sigma X^2 & &= 1662.72 \\
 (3) &= (61^2 + 27^2 + \dots + 55^2)/8 & &= 1956 \\
 (4) &= (40^2 + 53^2 + \dots + 36^2 + 49^2)/9 & &= 1844.75 \\
 (5) &= [(-3)^2 + 0^2 + \dots + 12^2 + 3^2 + \dots + (-18)^2]/108 & &= 1710.44 \\
 (6) &= [0^2 + (-6)^2 + \dots + 3^2 + (-9)^2 + \dots + 24^2]/54 & &= 13.81 \\
 (7) &= [1^2 + (-8)^2 + \dots + 0^2 + (-13)^2 + \dots + 38^2]/108 & &= 34.19 \\
 (8) &= [(-2)^2 + (-2)^2 + \dots + 9^2 + (-1)^2 + \dots + (-4)^2]/54 & &= 32.48 \\
 & & &= 15.52
 \end{aligned}$$

Source	SS	df	MS
Replications	(4) - (1) = 47.72	7	
Treatments (unadj)	(3) - (1) = 182.03	8	22.75
Residual	(2) - (3) - (4) + (1) = 63.53	56	1.13
Total	(2) - (1) = 293.28	71	

intrablock error. Since the treatment effects are not orthogonal to the row and column effects, there is not complete additivity for the adjusted variation.

The subdivision of the residual is given in Table 9.4-6. Under one subdivision, an estimate of the column variation (adjusted for treatments and rows) is obtained. Under a second subdivision, an estimate of the row variation (adjusted for treatments and columns) is obtained. Both subdivisions provide the same estimate of the intrablock variation; within rounding error the entries should be identical. The expected values of the

corresponding mean squares are as follows:

$$E(E_r) = \sigma_e^2 + \frac{nk - 1}{n} \sigma_{rows}^2,$$

$$E(E_c) = \sigma_e^2 + \frac{nk - 1}{n} \sigma_{columns}^2,$$

$$E(E_e) = \sigma_e^2.$$

The weighting factors needed to obtain adjusted treatment totals are also defined in Table 9.4-6. In terms of expected values of the mean squares, it may be shown that

$$E(W_r) = \frac{1}{\sigma_e^2 + k\sigma_{rows}^2},$$

$$E(W_c) = \frac{1}{\sigma_e^2 + k\sigma_{columns}^2}$$

$$E(W_e) = \frac{1}{\sigma_e^2}.$$

Adjusted treatment totals have the general form

$$T'_i = T_i + \lambda' L_i + \mu' M_i,$$

where λ' and μ' are defined in Table 9.4-6. The effective error associated with differences between adjusted totals is

$$\begin{aligned} E'_e &= E_e[1 + k(\lambda' + \mu')] = .647[1 + 3(.0429 + .0846)] \\ &= .894. \end{aligned}$$

Adjusted treatment totals and means are summarized in Table 9.4-7. These adjusted totals combine interrow and intercolumn treatment information with the intrarow and intracolumn treatment information.

A test on the hypothesis that $\tau_1 = \tau_3$, for example, uses the statistic

$$\begin{aligned} F &= \frac{(T'_1 - T'_3)^2}{2n(k+1)E'_e} = \frac{(61.47 - 40.55)^2}{4(4)(.894)} \\ &= 30.60. \end{aligned}$$

A critical value, for significance level .05 under the Scheffé procedure, is $8F_{.95}(8,24) = 18.9$. In summarizing the information on adjusted means, any of the test procedures described in Sec. 3.7 may be used.

An approximate over-all test on the hypothesis of no treatment differences is given by

$$F = \frac{MS_{T'}}{E'_e},$$

where

$$MS_{T'} = \frac{[\sum(T'_i)^2/n(k+1)] - (1)}{k^2 - 1}.$$

Table 9.4-6 Computation of Error Terms

Source	SS	df	MS
<u>Residual</u>			
Rows w. rep (adj for treat)	$(2) - (3) - (4) + (1) = 63.53$	56	
Col. w. rep (adj for treat and rows)	$(5) = 13.81$	16	
Intrablock error	$(6) = 34.19$	16	$2.137 = E_e$
	$(2) - (3) - (4) + (1)$ $- (5) - (6) = 15.53$	24	$.647 = E_e$
<u>Residual</u>			
Col. w. rep (adj for treat)	63.53	56	
Rows w. rep (adj for treat and col)	$(7) = 32.48$	16	
Intrablock error	$(8) = 15.52$	16	$.969 = E_r$
	$(2) - (3) - (4) + (1)$ $- (7) - (8) = 15.53$	24	$.647 = E_e$
$W_e = \frac{1}{E_e} = 1.546$ $W_r = \frac{nk - 1}{nkE_r - E_e} = \frac{nk - 1}{nkE_e - E_e} = \frac{5}{6(2.137) - (.647)} = .411$ $\lambda' = \frac{W_e - W_r}{k[W_r + W_e + (k - 1)W_e]} = \frac{.576}{3(4.471)} = .0429$ $\mu' = \frac{W_e - W_c}{k[W_r + W_e + (k - 1)W_e]} = \frac{1.135}{3(4.471)} = .0846$			

In this case,

$$MS_{T'} = \frac{1877.02 - 1662.72}{8} = \frac{214.30}{8} = 26.79,$$

and
$$F = \frac{26.79}{.894} = 29.97.$$

For a .05-level test, the critical value is $F_{.95}(8,24) = 2.36$.

Table 9.4-7 Adjusted Treatment Totals and Means

Treatment	$T'_i = T_i + (.0429)L'_i + (.0846)M'_i$	\bar{T}'_i	\bar{T}_i
1	61.47	7.7	7.6
2	25.35	2.9	3.4
3	40.55	5.1	4.8
4	28.36	3.5	4.0
5	47.95	6.0	6.1
6	28.74	3.6	3.4
7	34.67	4.3	4.2
8	20.67	2.6	2.9
9	58.24	7.3	6.9
	$G = 346.00$	43.0	43.3
		$G/n(k+1) = 43.2$	

9.5 Balanced Incomplete-block Designs

The balanced lattice designs considered in Sec. 9.3 are special cases of the general class of balanced incomplete-block designs. In the general case the following notation is used:

t = number of treatments.

k = block size.

b = number of blocks.

r = number of replications.

λ = number of times each pair of treatments occurs together within some block.

A necessary condition for a balanced design is that each possible pair of treatments occurs together within some block the same number of times; that is, λ is a constant for all pairs.

The design given in Table 9.5-1 is balanced in this sense. It will be noted that treatment 1 is paired with treatment 2 in blocks 1, 2, and 6. It will also be noted that treatment 1 is paired with every other treatment in three blocks. Hence $\lambda = 3$ for treatment 1. Similarly, $\lambda = 3$ for all the other treatments. Inspection of this design will show that each of the treatments occurs in 6 of the 10 blocks; hence there are six replications. In this particular design, the blocks cannot be grouped into distinct replications. However, blocks 1 through 5 include three replications, and blocks 6 through 10 include

three replications. In some balanced incomplete-block designs it is possible to group the blocks into distinct replications. In the design given in Table 9.5-1, $b = 10$. A question might be raised about the minimum number of blocks required for balance, assuming t and k specified.

Table 9.5-1 Balanced Incomplete-block Design
($t = 5$, $k = 3$, $b = 10$, $r = 6$, $\lambda = 3$)

Block	Treatments			Block	Treatments		
(1)	1	2	3	(6)	1	2	4
(2)	1	2	5	(7)	1	3	4
(3)	1	4	5	(8)	1	3	5
(4)	2	3	4	(9)	2	3	5
(5)	3	4	5	(10)	2	4	5

The total number of observations in an experiment having b blocks, with k observations in each block, is bk . If there are r replications of t treatments, then the total number of observations is rt . Thus, the following relation must hold:

$$bk = rt.$$

From this relation,

$$b = \frac{rt}{k}.$$

If $k = 3$ and $t = 5$,

$$b = \frac{5r}{3}.$$

Since b must be a whole number, $r = 3$ is the smallest number which will make b equal to a whole number. With $r = 3$ and $b = 5$, it is not possible to construct a balanced design for the specified k and t in five blocks. The next larger value of r that will make b a whole number is 6. When $r = 6$ and $b = 10$, it is possible to construct a balanced design having 10 blocks; further, this is the minimum number of blocks for which a balanced design (having $t = 5$ and $k = 3$) may be constructed. The minimum block size for other combinations of t and k is given in Table 9.5-2. Balanced incomplete-block designs are available for these combinations of t and k with the minimum b indicated. A more extensive table of available balanced incomplete-block designs will be found in Cochran and Cox (1957, pp. 469-470).

With block size k , treatment i may be paired with $k - 1$ other treatments within a single block. If there are r replications, treatment i will appear in r different blocks. Hence the total number of pairs in which treatment i appears within some common block will be $r(k - 1)$. For example, in the

design in Table 9.5-1 treatment 1 occurs in two pairs with each of blocks 1, 2, 3, 6, 7, and 8. Hence the total number of pairs occurring within some common block which include treatment 1 is $6(2) = 12$.

In a balanced design each treatment is paired within some common block with the other $t - 1$ treatment λ times. Hence the total number of pairs occurring within some common block must be equal to $\lambda(t - 1)$ if the design

Table 9.5-2 Some Available Balanced Incomplete-block Designs

t	k	r	b	E
4	2	3	6	.67
5	2	4	10	.62
5	3	6	10	.83
6	2	5	15	.60
6	3	5	10	.80
6	4	10	15	.90
7	2	6	21	.58
7	3	3	7	.78
7	4	4	7	.88
8	2	7	28	.57
8	4	7	14	.86
9	2	8	36	.56
9	3	4	12	.75
9	4	8	18	.84
9	5	10	18	.90
10	2	9	45	.56
10	3	9	30	.74
10	4	6	15	.83
10	5	9	18	.89
10	6	9	15	.93

is to be balanced. In general, for a balanced incomplete-block design the following relation must hold:

$$r(k - 1) = \lambda(t - 1).$$

By definition, the efficiency factor, designated by the symbol E , is

$$E = \frac{t(k - 1)}{k(t - 1)}.$$

It will be found that the number of effective replications for within-block treatment information will be equal to rE . By using the necessary relationships for a balanced design, through straightforward algebraic manipulation it may be shown that

$$krE = t\lambda = kr - r + \lambda.$$

The analysis of variance for the case in which the replication dimension is disregarded has the following form:

Source	df	MS
Treatments (unadj)	$t - 1$	E_b E_e
Blocks (adj)	$b - 1$	
Intrablock error	$rt - t - b + 1$	
Total	$rt - 1$	

Before indicating the method for obtaining the sums of squares in the analysis of variance, some of the principles underlying the adjustments will be indicated. These principles are identical to those used in balanced lattice designs.

Assuming a linear model in which block effects are strictly additive, the sum of the r observations on treatment i provides an estimate of the parameters on the right-hand side of the following expression:

$$T_i \doteq r\mu + r\tau_i + \sum_{(i)} \beta_j,$$

where the last symbol on the right-hand side is the sum of the block effects in which treatment i appears. Under the same model, the sum of the block totals in which treatment i appears estimates the parameters on the right-hand side of the following expression:

$$B_{(i)} \doteq kr\mu + r\tau_i + \lambda \sum \tau_{i'} + k \sum_{(i)} \beta_j,$$

where $\tau_{i'}$ represents any treatment except τ_i . Since $\sum \tau_i = 0$, it follows that

$$\begin{aligned} r\tau_i + \lambda \sum \tau_{i'} &= (r - \lambda)\tau_i + \lambda \tau_i + \lambda \sum \tau_{i'} \\ &= (r - \lambda)\tau_i. \end{aligned}$$

This is so because $\lambda \tau_i + \lambda \sum \tau_{i'} = \lambda \sum \tau_i = 0$. Hence,

$$B_{(i)} \doteq kr\mu + (r - \lambda)\tau_i + k \sum_{(i)} \beta_j.$$

In terms of T_i and $B_{(i)}$, an unbiased estimate of the effect of treatment i is given by

$$\begin{aligned} kT_i &\doteq kr\mu + kr\tau_i + k \sum_{(i)} \beta_j \\ B_{(i)} &\doteq kr\mu + (r - \lambda)\tau_i + k \sum_{(i)} \beta_j \\ \hline Q_i = kT_i - B_{(i)} &\doteq (kr - r + \lambda)\tau_i \\ &\doteq krE\tau_i \end{aligned}$$

If there were no confounding with block effects, the right-hand side of this last expression would be $kr\tau_i$. The efficiency factor E , which is always less than or equal to unity, is a measure of the relative amount of within-block

treatment information per replication. A balanced incomplete-block design is more precise than a randomized- (complete-)block design only if

$$\sigma_e^2 < E\sigma^2,$$

where σ_e^2 is the error variance for the incomplete-block design and σ^2 is the error variance for the randomized-block design having an equal number of replications.

The sums of squares required in the analysis of variance are obtained as follows:

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{rt},$$

$$SS_{\text{treat(unadj)}} = \frac{\Sigma T_i^2}{r} - \frac{G^2}{rt},$$

$$SS_{\text{treat(adj for blocks)}} = \frac{\Sigma Q_i^2}{k^2 r E},$$

$$SS_{\text{blocks(unadj)}} = \frac{\Sigma B_j^2}{k} - \frac{G^2}{rt},$$

$$SS_{\text{blocks(adj)}} = SS_{\text{blocks(unadj)}} + SS_{\text{treat(adj for blocks)}} - SS_{\text{treat(unadj)}},$$

$$SS_{\text{intrablock error}} = SS_{\text{total}} - SS_{\text{treat(unadj)}} - SS_{\text{blocks(adj)}}.$$

An estimate of a treatment total, which combines intrablock and inter-block information, is given by

$$T'_i = T_i + \mu W_i,$$

where

$$\mu = \frac{(b-1)(E_b - E_e)}{t(k-1)(b-1)E_b + (t-k)(b-t)E_e},$$

and

$$W_i = (t-k)T_i + (t-1)B_{(i)} + (k-1)G.$$

(If E_b is less than E_e , then the value of μ is taken to be zero.) It should be noted that the definition of W_i in this design differs from the corresponding definition for a lattice design. W_i in this case is equivalent to $k-1$ times W_i for the balanced lattice. The effective error for adjusted treatment totals is

$$E'_e = E_e[1 + (t-k)\mu].$$

A test on the difference between two adjusted treatment totals uses the statistic

$$F = \frac{(T'_i - T'_{i'})^2}{2rE'_e}.$$

If means are used instead of treatment totals, the F ratio is

$$F = \frac{(\bar{T}'_i - \bar{T}'_{i'})^2}{2(E'_e/r)}.$$

For a .05-level test the critical value is $F_{.95}[1, (rt - t - b + 1)]$. Any of the methods for making multiple comparisons discussed in Chap. 3 may be adapted for use in this analysis.

The over-all variation due to the adjusted treatment effects is

$$SS_{\text{treat(adj)}} = \frac{\sum(T_i')^2}{r} - \frac{G^2}{rt}.$$

This source of variation has $t - 1$ degrees of freedom. An approximate over-all test on differences between treatment effects is given by

$$F = \frac{MS_{\text{treat(adj)}}}{E'_e}.$$

If only the intrablock information on treatments is used, a test on differences between treatment effects is given by

$$F = \frac{MS_{\text{treat(adj for blocks)}}}{E_e}.$$

The latter F ratio will follow an F distribution when the usual analysis of variance assumptions are met. However, the F ratio having E'_e as a denominator does not quite follow the usual F distribution, since no account is taken of the sampling error in the estimation of the weighting factor μ . The sampling error in μ does not have a marked effect upon the distribution of the F ratio when the number of blocks is greater than 10.

For designs in which, simultaneously, k is less than 5 and b is less than 10, interblock information on treatments is generally disregarded. In this case, the adjusted treatment total is given by

$$T_i'' = \frac{Q_i}{kE} + r\bar{G}.$$

An adjusted treatment mean is given by

$$\bar{T}_i'' = \frac{Q_i}{krE} + \bar{G}.$$

Thus \bar{T}_i'' is a treatment mean based solely upon intrablock information, whereas \bar{T}_i' is a treatment mean which combines both intrablock and interblock treatment information. A test on the difference between two treatment means which is based only on intrablock information uses the following test statistic,

$$F = \frac{(T_i'' - T_{i'}'')^2}{2r(E_e/E)},$$

where $E = t(k - 1)/k(t - 1)$.

The adjusted sum of squares for blocks may be computed by an alternative method. As a first step one computes the treatment component of the blocks (ignoring treatments), which is given by

$$SS_1 = \frac{\sum B_{(i)}^2 - [(\sum B_{(i)}^*)^2/t]}{k(r - \lambda)}.$$

What is called the remainder sum of squares is given by

$$SS_{\text{rem}} = SS_{\text{blocks}} - SS_1.$$

The remainder sum of squares has $b - t$ degrees of freedom. The treatment component of the blocks (adjusted for treatments) is given by

$$SS_2 = \frac{\sum W_i^2}{rt(t - k)(k - 1)}.$$

This source of variation has $t - 1$ degrees of freedom. The adjusted sum of squares for blocks is given by

$$SS_{\text{blocks(adj)}} = SS_{\text{rem}} + SS_2,$$

with degrees of freedom

$$b - 1 = (b - t) + (t - 1).$$

In this latter form, the adjusted sum of squares for blocks is seen to have two components. The expected value of the mean squares due to the components as well as the pooled sources are as follows:

$$E(MS_{\text{rem}}) = \sigma_e^2 + k\sigma_\beta^2,$$

$$E(MS_2) = \sigma_e^2 + kE\sigma_\beta^2,$$

$$E(MS_{\text{blocks(adj)}}) = \sigma_e^2 + \frac{bk - t}{b - 1} \sigma_\beta^2.$$

Designs Arranged in Distinct Replications. Some of the balanced incomplete-block designs may be arranged in distinct replications. The analysis of variance for such designs, provided that the experiment is conducted as a replicated experiment, has the following form:

Source	df	MS
Replications	$r - 1$	
Treatments (unadj)	$t - 1$	
Blocks within reps (adj)	$b - r$	E_b
Intrablock error	$rt - t - b + 1$	E_e
Total	$rt - 1$	

The analysis of this design is identical to that which has just been outlined, except for two changes introduced by the fact that the blocks are nested within the replications.

The variation due to replications is given by

$$SS_{\text{reps}} = \frac{\sum R_m^2}{t} - \frac{G^2}{rt},$$

where R_m is the sum of the t observations in replication m . Since the blocks are nested, $SS_{\text{blocks within reps}}$ replaces SS_{blocks} throughout.

$$SS_{\text{Blocks within reps(unadj)}} = \frac{\sum B_j^2}{k} - \frac{\sum R_m^2}{t}.$$

The adjusted variation due to blocks within replication is given by the following relation:

$$SS_{\text{blocks w. reps (adj)}} = SS_{\text{blocks w. reps}} + SS_{\text{treat (adj)}} - SS_{\text{treat (unadj)}}.$$

Computation of the treatment variation is identical to that for the case in which there are no distinct replications.

The definition of μ for this type of design is different from that given previously. In this case,

$$\mu = \frac{r(E_b - E_e)}{rt(k-1)E_b + k(b-r-t+1)E_e}.$$

The rest of the analysis has the same pattern as the design which has just been considered.

Repetitions of Designs Not Having Distinct Replications. A balanced incomplete-block design will provide r observations on each treatment. If it is desired to have nr observations rather than r , the design may be repeated n times, with independent randomizations each time. It is convenient to let $b = b'n$, where b' is the number of blocks in the original design. It will also be convenient to let $r = r'n$, where r' is the number of replications in the original design. Thus, in this context, b is the number of blocks in the enlarged plan, and r is the number of replications in the enlarged plan. In terms of this notation, the analysis of variance has the following form:

Source	df	MS
Repetitions	$n - 1$	
Treatments (unadj)	$t - 1$	
Blocks within repetition (adj)	$b - n$	E_b
Intrablock error	$rt - t - b + 1$	E_e
Total	$nrt - 1$	

The sums of squares are obtained as follows:

$$(i) = SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{rt}.$$

$$(ii) = SS_{\text{repetitions}} = \frac{\Sigma (\text{Rep})^2}{rt/n} - \frac{G^2}{rt}.$$

$$(iii) = SS_{\text{treat (unadj)}} = \frac{\Sigma T_i^2}{r} - \frac{G^2}{rt}.$$

$$(iv) = SS_{\text{treat (adj for blocks)}} = \frac{\Sigma Q_i^2}{k^2 r E}.$$

$$(v) = SS_{\text{blocks w. repetitions}} = \frac{\Sigma B_j^2}{k} - \frac{\Sigma (\text{Rep})^2}{rt/n}.$$

The adjusted sum of squares for blocks within repetitions is given by

$$(vi) = SS_{\text{blocks w. repetitions(adj)}} = (iv) + (v) - (iii).$$

The intrablock error is

$$SS_{\text{intrablock error}} = (i) - (ii) - (iii) - (vi).$$

For this case,

$$\mu = \frac{(b-n)(E_b - E_e)}{t(k-1)(b-n)E_b + (t-k)(b-t-n+1)E_e}.$$

(It will be noted that when $n = 1$ this definition of μ reduces to that given for the first design in this section.) The procedure for obtaining adjusted treatment totals and making tests is analogous to that followed for a single run of the basic balanced design.

If the block corresponds to a person, this type of design calls for n groups of subjects with b' subjects per group. The subjects within a group are assigned at random to the blocks defined by the basic plan. Each subject is given the k treatments within the block to which he is assigned; order of administration is randomized independently for each subject. In terms of a repeated-measure design with the subjects corresponding to blocks, the analysis of variance has the following form:

Source	df	MS
Groups	$n - 1$	
Treatments (unadj)	$t - 1$	
Subj w. groups (adj)	$b - n$	E_b
Within-subject error	$rt - t - b + 1$	E_e

If E_b is large relative to E_e , only within-block information on treatments is required. If between-block treatment information is disregarded, the analysis of variance takes the following form:

Source	df	MS
Groups	$n - 1$	
Subj w. groups (unadj)	$b - n$	
Treatments (adj for blocks)	$t - 1$	
Within-block error	$rt - t - b + 1$	E_e

The adjusted sum of squares for treatments is given by

$$SS_{\text{treat(adj for blocks)}} = \frac{\sum Q_i^2}{k^2 r E}.$$

It is readily verified that

$$SS_{\text{subj w. group(unadj)}} + SS_{\text{treat(adj for blocks)}} = SS_{\text{subj w. group(adj)}} + SS_{\text{treat(unadj)}}.$$

When between-block treatment information is disregarded, E_e is the proper error mean square in statistical tests. In testing differences between means, each mean is considered as being based upon rE replications rather than r replications. Thus the standard error of a difference is $\sqrt{E_e/(rE)}$ instead of $\sqrt{E_e/r}$.

9.6 Numerical Example of Balanced Incomplete-block Design

Computational procedures will be illustrated by means of the data in Table 9.6-1. In the experiment from which these data were obtained, there are $t = 6$ treatments arranged in blocks of size $k = 3$. A balanced design

Table 9.6-1 Numerical Example
($t = 6, k = 3, r = 5, b = 10, \lambda = 2, E = .80$)

Subject				Total	Subject				Total
1	(1) 2	(2) 3	(5) 8	13	6	(2) 3	(3) 6	(4) 5	14
2	(1) 5	(2) 4	(6) 10	19	7	(2) 4	(3) 7	(5) 10	21
3	(1) 4	(3) 8	(4) 8	20	8	(2) 7	(4) 7	(6) 11	25
4	(1) 7	(3) 9	(6) 12	28	9	(3) 9	(5) 12	(6) 11	32
5	(1) 6	(4) 9	(5) 13	28	10	(4) 9	(5) 14	(6) 13	36

	T_i	$B_{(i)}$	$Q_i = 3T_i - B_{(i)}$	$W_i = 3T_i - 5B_{(i)} + 2G$	$T'_i = T_i + .078W_i$
1	24	108	-36	4	24.3
2	21	92	-29	75	26.8
3	39	115	2	14	40.1
4	38	123	-9	-29	35.7
5	57	130	41	-7	56.5
6	57	140	31	-57	52.6
	236 = G	708 = 3G	0	0	236.0

requires a minimum of 10 blocks. This design corresponds to plan 11.4 in Cochran and Cox (1957, p. 471). There are $r = 5$ replications, but the blocks cannot be arranged in distinct replications. Each treatment is paired with each of the other treatments within a common block $\lambda = 2$ times.

Subjects are assigned at random to the blocks. For example, subject 6 is assigned to the block having treatments 2, 3, and 4. The criterion scores on subject 6 are, respectively, 3, 6, and 5. The sequences in which the treatments are given the subjects are randomized independently. In this

case there are $k = 3$ measures on each subject, the measures being the criterion scores on the treatments assigned to the blocks. The sum of the criterion scores for an individual j is B_j . For example,

$$B_6 = 3 + 6 + 5 = 14.$$

Sums required in the analysis of variance are given in the lower half of Table 9.6-1. An entry in column T_i represents the sum of the $r = 5$ criterion scores on treatment i . For example, T_3 is obtained as follows:

*Criterion
score*

8	from subject 3
9	from subject 4
6	from subject 6
7	from subject 7
9	from subject 9
<hr/> 39 = T_3	

An entry in column $B_{(i)}$ is the sum of the $r = 5$ totals for blocks in which treatment i appears. For example, $B_{(3)}$ is obtained as follows:

B_j	
20	from subject 3
28	from subject 4
14	from subject 6
21	from subject 7
32	from subject 9
<hr/> 115 = $B_{(3)}$	

The formulas for Q_i and W_i are given at the top of their respective columns. For example,

$$Q_3 = 3(39) - 115 = 2,$$

$$W_3 = 3(39) - 5(115) + 2(236) = 14.$$

The general formula for T'_i is

$$T'_i = T_i + \mu W_i.$$

Hence the value of μ must be obtained before T'_i can be computed. The value of μ is obtained after the preliminary analysis is completed.

Convenient computational symbols are summarized in the upper part of Table 9.6-2. Symbol (2) is obtained from the individual criterion scores; data for all other symbols are in the lower half of Table 9.6-1. A partial check on the computational symbols is indicated. This check utilizes the fact that the remainder component of the sum of squares for blocks can be computed in two different ways.

In terms of these computational symbols, formulas for sums of squares in the analysis of variance are given in the lower portion of Table 9.6-2. The analysis of variance is summarized in Table 9.6-3. The weighting factor μ and the effective mean square for error E_e are also given in Table 9.6-3.

Table 9.6-2 Computation of Sums of Squares

(1) = $G^2/(rt)$ = 1856.53	(4) = $(\Sigma B_j^2)/k$ = 2026.67
(2) = ΣX^2 = 2168	(5) = $(\Sigma Q_i^2)/(k^2rE)$ = 135.11
(3) = $(\Sigma T_i^2)/r$ = 2096.00	(6) = $\frac{\Sigma W_i^2}{rt(t-k)(k-1)}$ = 55.42
(7) = $\frac{\Sigma B_{(i)}^2 - [(\Sigma B_{(i)})^2/t]}{k(r-\lambda)}$ = $\frac{84,982 - 83,544}{9}$ = 159.78	

Check: (4) - (1) - (7) = (4) + (5) - (3) - (6)
 $10.36 = 10.36$

SS _{total}	= (2) - (1)	= 311.47
SS _{treat(unadj)}	= (3) - (1)	= 239.47
SS _{blocks(unadj)}	= (4) - (1)	= 170.14
SS _{blocks(adj)}	= (4) + (5) - (3)	= 65.78
SS _{intra-block error}	= (2) - (4) - (5)	= 6.22

Table 9.6-3 Summary of Analysis of Variance

Source	SS	df	MS
Treatments (unadj)	239.47	5	
Blocks (adj)	65.78	9	7.31 = E_b
Intra-block error	6.22	15	.42 = E_e
Total	311.47	29	

$$\mu = \frac{(b-1)(E_b - E_e)}{t(k-1)(b-1)E_b + (t-k)(b-t)E_e} = \frac{9(7.31 - .42)}{6(2)(9)(7.31) + (3)(4)(.42)}$$

$$= .078$$

$$E'_e = E_e[1 + (t-k)\mu] = .42[1 + 3(.078)] = .52$$

Table 9.6-4 Adjusted Treatment Means

Treatment	$T'_i = T'_i/r$	$T''_i = \frac{Q_i}{krE} + \bar{G}$
1	4.86	4.87
2	5.36	5.45
3	8.02	8.04
4	7.14	7.12
5	11.30	11.29
6	10.52	10.45
	47.20	47.22

$$t\bar{G} = 6(7.87) = 47.22$$

Given the value of μ , entries in the column headed T'_i in Table 9.6-1 may now be computed. The latter entries are the adjusted treatment totals for the combined intrablock and interblock data.

Adjusted treatment means are summarized in Table 9.6-4. Two sets of means are given. \bar{T}'_i includes intra- as well as interblock information, whereas \bar{T}''_i is based solely on intrablock information. In this case the two sets of means differ only in the second decimal place. Whenever E_b is large relative to E_e (in this case $E_b/E_e = 7.31/.42 = 17.4$), the contribution of the interblock information relative to the intrablock information will be relatively small. Thus, when E_b is large relative to E_e ,

$$\bar{T}'_i \doteq \bar{T}''_i.$$

The over-all variation between the adjusted treatment totals is

$$\begin{aligned} \text{SS}_{\text{treat(adj)}} &= \frac{\Sigma(T'_i)^2}{r} - \frac{G^2}{rt} \\ &= 2030.05 - 1856.53 = 173.52. \end{aligned}$$

The corresponding mean square is $173.52/(t-1) = 34.70$. An approximate test of the hypothesis that there are no differences between the treatments, that is, $\sigma_\tau^2 = 0$, uses the statistic

$$F = \frac{\text{MS}_{\text{treat(adj)}}}{E'_e} = \frac{34.70}{.52} = 66.73.$$

The critical value for a .01-level test is $F_{.99}(5,15) = 4.56$. The test indicates that the treatment effects do differ at the .01 level of significance.

A test of the hypothesis that $\tau_3 = \tau_4$, against a two-tailed alternative hypothesis, uses the following statistic (these two treatment effects are used for illustrative purposes):

$$F = \frac{(T'_3 - T'_4)^2}{2rE'_e} = \frac{(40.1 - 35.7)^2}{2(5)(0.52)} = 4.56.$$

If this comparison may be considered as being in the a priori category, the critical value for a .01-level test is $F_{.99}(1,15) = 8.68$. If this comparison is one among a relatively large number of comparisons that are to be made, and if an α level for the whole collection of tests is desired, the critical value (using the Scheffé approach) is

$$(t-1)F_{.99}[(t-1), df_{E'_e}] = 5F_{.99}[5,15] = 5(4.56) = 22.80.$$

The Newman-Keuls procedure for testing differences between ordered pairs of adjusted means is illustrated in Table 9.6-5. (In making these tests it is more convenient to work with the adjusted totals rather than the adjusted means.) In this table, m is the number of steps two means are apart.

When only the intrablock information is used in estimating treatment effects, E_e/E replaces E'_e in making comparisons. Since the efficiency factor

E is always less than unity, E_e/E will be greater than E_e . In the numerical example under consideration,

$$E_e/E = .42/.80 = .54.$$

The smaller the efficiency factor, the greater is the difference between E_e and E_e/E , which is the effective error for comparisons for tests which use only intrablock information.

Table 9.6-5 Tests on Differences between All Pairs of Treatments (Newman-Keuls)

Ordered T'_i :	(1)	(2)	(4)	(3)	(6)	(5)
	24.3	26.8	35.7	40.1	52.5	56.5
Differences between all pairs	(1)	(2)	(4)	(3)	(6)	(5)
		2.5	11.4*	15.8*	28.2*	32.2*
	(2)		8.9*	13.3*	25.7*	29.7*
	(4)			4.4	16.8*	20.8*
	(3)				12.4*	16.4*
	(6)					4.0
$s_{T'} = \sqrt{rE'_e} = \sqrt{5(.52)} = 1.61$						
	$q_{.99}(m,15)$	4.17	4.83	5.25	5.56	5.80
	$(1.61)q_{.99}(m,15)$	6.71	7.78	8.45	8.95	9.34
						Critical values

It is of interest to point out some of the relationships that exist between the sums of squares in a balanced incomplete-block design. For example,

$$SS_{\text{blocks(unadj)}} + SS_{\text{treat(adj for blocks)}} = SS_{\text{blocks(adj)}} + SS_{\text{treat(unadj)}},$$

$$170.14 + 135.11 = 65.78 + 239.47,$$

$$305.25 = 305.25.$$

Also,

$$SS_{\text{rem}} = SS_{\text{blocks(unadj)}} - SS_1$$

$$= 170.14 - 159.78$$

$$= 10.36,$$

$$SS_{\text{rem}} = SS_{\text{blocks(adj)}} - SS_2$$

$$= 65.78 - 55.42$$

$$= 10.36.$$

In the above expressions,

$$SS_2 = \frac{\sum W_i^2}{rt(t-k)(k-1)},$$

$$SS_1 = \frac{\sum B_{(i)}^2 - [(\sum B_{(i)})^2/t]}{k(r-\lambda)}.$$

Estimation of Missing Data. One method of estimating a missing cell entry will be illustrated by means of the numerical example. The principle underlying this method of estimation is that of minimizing the intrablock error in the subsequent analysis of variance. Suppose, in Table 9.6-1, that the entry under treatment 6 for subject 4 were missing. If an x were substituted for the missing entry, the totals in the lower part of this table would have the following form:

Treatment	T_i	$B_{(i)}$	Q_i
1	24	$96 + x$	$-24 - x$
2	21	92	-29
3	39	$103 + x$	$14 - x$
4	38	123	-9
5	57	140	41
6	$45 + x$	$128 + x$	$7 + 2x$

If B_j = sum of observations present in block with missing entry,

Q_i = Q value for treatment which has missing entry,

Q'_i = sum of Q values for treatments in same block as missing entry,
then an estimate of the missing entry is given by

$$x_{ij} = \frac{rt(k-1)B_j + k(t-1)Q_i - (t-1)Q'_i}{(k-1)[rt(k-1) - k(t-1)]}$$

For the case being considered, the missing cell entry may be designated x_{64} , that is, an observation on treatment 6 in block 4.

$$B_4 = 7 + 9 = 16,$$

the observations in block 4,

$$Q_6 = 7,$$

disregarding x ,

$$Q'_i = Q_1 + Q_3 = -24 + 14 = -10$$

disregarding x .

Q' has the value indicated since treatments 1 and 3 are present in the block with the missing cell entry. Thus

$$\begin{aligned} x_{64} &= \frac{30(2)(16) + (3)(5)(7) - (5)(-10)}{2[(30)(2) - 3(5)]} \\ &= \frac{1115}{90} = 12.4. \end{aligned}$$

Estimation procedures for the case of more than a single missing entry are illustrated in Cochran and Cox (1957, pp. 451-452). For each missing entry estimated one degree of freedom is subtracted from the intrablock error. Once the missing entries are estimated, the analysis proceeds in the usual way.

In well-planned experimental work, particularly in cases in which a subject corresponds to a block, the possibility of missing data can be avoided

by having reserve experimental units available for use should the need arise. In many cases such needs do arise. Estimation procedures for missing cell entries are at the very best only poor substitutes for the actual observations.

9.7 Youden Squares

In a balanced lattice-square design there is balance with respect to both row and column. An analogous balance applied to a balanced incomplete-block design is achieved in designs known as Youden squares. Although the shape of a Youden square is a rectangle, the name stems from the original method of its construction—namely, by deleting certain columns of special kinds (cyclic) of Latin squares. W. J. Youden developed and made extensive use of such designs in agricultural experimentation. Some interesting applications of Youden squares to psychological research are discussed by Durbin (1951).

An example of a Youden square is given in Table 9.7-1. With respect to

Table 9.7-1 Youden Square
($t = b = 7, k = r = 3, \lambda = 1, E = .78$)

Block	Column		
	1	2	3
1	4	6	7
2	5	7	1
3	6	1	2
4	7	2	3
5	1	3	4
6	2	4	5
7	3	5	6

the blocks (rows), this design has all the properties of a balanced incomplete-block design. In addition, there is balance with respect to the columns—each treatment appears once and only once within each of the columns. Hence each column of a Youden square forms a distinct replication. It will be found, however, that the blocks cannot be grouped into distinct replications.

In general, for a Youden square, $t = b$, and $k = r$. In cases in which the block represents a subject in a repeated-measure design, the columns may represent the order in which the treatments are administered to a subject. The analysis of variance for a Youden square follows the same general pattern as that given for the case of a balanced incomplete-block design in which the blocks cannot be arranged into distinct replications. Because $t = b$, the analysis becomes greatly simplified.

The analysis of variance for a Youden square has the following form:

Source	df	MS
Columns	$k - 1$	
Treatments (unadj)	$t - 1$	
Blocks (adj)	$t - 1$	E_b
Intrablock error	$(k - 2)(t - 2)$	E_e
	$kt - 1$	

The sums of squares are computed as follows (symbols have the same definitions as those used for the balanced incomplete-block design):

$$SS_{\text{total}} = \Sigma X^2 - \frac{G^2}{kt},$$

$$SS_{\text{columns}} = \frac{\Sigma(\text{col total})^2}{t} - \frac{G^2}{kt},$$

$$SS_{\text{treat(unadj)}} = \frac{\Sigma T_i^2}{t} - \frac{G^2}{kt},$$

$$SS_{\text{blocks(adj)}} = \frac{\Sigma W_i^2}{rt(t - k)(k - 1)},$$

$$SS_{\text{intrablock error}} = SS_{\text{total}} - SS_{\text{col}} - SS_{\text{treat(unadj)}} - SS_{\text{blocks(adj)}}.$$

Because $b = t$, there is only one component, the treatment component, of the adjusted sum of squares for blocks; hence the simplified computation of $SS_{\text{blocks(adj)}}$ as compared with the general case of a balanced incomplete-block design.

The weighting factor for a Youden square simplifies to

$$\mu = \frac{E_b - E_e}{t(k - 1)E_b}.$$

An adjusted treatment total has the form

$$T'_i = T_i + \mu W_i.$$

A test on the difference between two treatment effects has the general form

$$F = \frac{(T'_i - T'_{i'})^2}{2kE'_e},$$

where

$$E'_e = [1 + (t - k)\mu].$$

In the analysis outlined above, both intrablock and interblock treatment information are used. If only intrablock information on treatments is

used, the analysis of variance has the following form:

Source	df	MS
Columns	$k - 1$	E_e
Blocks (unadj)	$t - 1$	
Treatments (adj for blocks)	$t - 1$	
Intrablock error	$(k - 2)(t - 2)$	
	$kt - 1$	

The numerical value of E_e will be found to be the same in both analyses. The sum of squares for treatments adjusted for blocks is

$$SS_{\text{treat(adj for blocks)}} = \frac{\sum Q_i^2}{kt}.$$

A treatment total, adjusted for blocks, is

$$T_i'' = \frac{Q_i}{kE} + k\bar{G}.$$

In this case a test on the difference between two treatment effects has the general form

$$F = \frac{(T_i'' - T_j'')^2}{2kE_e''},$$

where $E_e'' = E_e/E$.

Construction of Youden Squares. Cyclic Latin squares are defined in Sec. 10.2. The following 5×5 Latin square represents a cyclic Latin square:

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>
1	2	3	4	5
2	3	4	5	1
3	4	5	1	2
4	5	1	2	3
5	1	2	3	4

To construct a Youden square in which $t = 5$ and $k = 4$, any four of the five columns of this square will provide the required design.

The following 7×7 Latin square also represents a cyclic square:

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>	<i>e</i>	<i>f</i>	<i>g</i>
1	2	3	4	5	6	7
2	3	4	5	6	7	1
3	4	5	6	7	1	2
4	5	6	7	1	2	3
5	6	7	1	2	3	4
6	7	1	2	3	4	5
7	1	2	3	4	5	6

A Youden square in which $t = 7$ and $k = 3$ may be constructed from this square by selecting any two adjacent columns and a third column which is one step from being adjacent. For example, columns a , b , and d meet these requirements. Columns c , e , and f also meet these requirements as do columns d , e , and g . The Youden square given by columns d , e , and g is shown below:

Block	d	e	g
1	4	6	7
2	5	7	1
3	6	1	2
4	7	2	3
5	1	3	4
6	2	4	5
7	3	5	6

In using a design constructed in this manner, treatments are assigned at random to the numbers, and assignment of the blocks is also randomized. Columns within the blocks are likewise randomized.

Since Youden squares are special cases of balanced incomplete-block designs, they exist only where there is a balanced incomplete-block design in which $k = r$ and $t = b$.

A partial listing of some of the available Youden squares is given in Table 9.7-2. A more complete listing, as well as the structure of the corresponding plans, will be found in Cochran and Cox (1957, pp. 520-544).

Table 9.7-2 Index to Some Youden Squares

$t = b$	$k = r$	λ	E
5	4	3	.94
6	5	4	.96
7	3	1	.78
7	4	2	.88
8	7	6	.98
11	5	2	.88
11	6	3	.92
13	4	1	.81
13	9	6	.95
15	7	3	.92
15	8	4	.94
16	6	2	.89
16	10	6	.96

Repetitions of a Youden Square. A single Youden square provides k observations on each treatment. In order to increase the number of observations per treatment from k to nk , a Youden square may be repeated n

times. The columns of the Youden square should be randomized independently for each of the repetitions. The analysis of variance has the following general form:

Source	df	MS
Columns	$kn - 1$	E_b E_e
Treat (unadj)	$t - 1$	
Blocks w. reps (adj)	$n(t - 1)$	
Intrablock error	$(t - 1)(kn - n - 1)$	
Total	$nkt - 1$	

Computational procedures for this case are illustrated by the numerical example given in the next section.

9.8 Numerical Example of Youden Square

The data in Table 9.8-1 will be used to illustrate the computational procedures for the case of n repetitions of a Youden square. The basic plan has $t = b = 7$ treatments arranged in seven blocks. There are $k = r = 3$ treatments in each block, and there are three replications in the basic plan. Two repetitions of the basic plan are to be used. The columns have been independently randomized. The treatments are enclosed in parentheses. Corresponding blocks appear in corresponding positions on the left and right.

Suppose that $n = 2$ samples of $t = 7$ subjects each are drawn from a population of interest. The subjects in the first sample are assigned at random to the blocks in repetition 1; the subjects in the second sample are assigned at random to the blocks in repetition 2. In Table 9.8-1, subjects 1 and 8 are assigned to the same basic block; similarly, subjects 2 and 9 are assigned to the same basic block. Subject 1 is given treatments (4), (6), and (7) in this sequence. Subject 8 is given the same treatments, but in the sequence (7), (6), and (4). Similarly, subjects 2 and 9 are given the same set of treatments, but the sequence of administration differs.

Suppose that the "treatments" represent seven different packages which are to be assigned a rating on a nine-point scale of attractiveness. This rating defines the criterion score. For example, subject 1 assigned the ratings 2, 4, and 6, respectively, to packages (4), (6), and (7). The sum of the three ratings is 12.

The sum of the $k = 3$ observations on a subject corresponds to a block total. This sum will be designated by the symbol P_m ; thus P_1 is the sum of observations on subject 1. These block totals are summarized in the middle part of Table 9.8-1. There is a set of block totals for each repetition of the basic plan. Corresponding to each block in the basic plan there is also a set of n totals. The $n = 2$ totals for block 1 of the basic plan are 12 and 7. The sum of these totals, which is 19, will be designated by B_1 . Similarly,

Table 9.8-1 Data for Numerical Example

 $(t = 7, k = 3, n = 2, \lambda = 1)$

Repetition 1					Repetition 2				
Subject				Total	Subject				Total
1	(4) 2	(6) 4	(7) 6	$12 = P_1$	8	(7) 4	(6) 2	(4) 1	$7 = P_8$
2	(5) 6	(7) 9	(1) 3		9	(1) 5	(7) 9	(5) 3	
3	(6) 1	(1) 2	(2) 3	6	10	(2) 4	(1) 3	(6) 6	13
4	(7) 7	(2) 1	(3) 4	12	11	(3) 1	(2) 1	(7) 4	6
5	(1) 1	(3) 1	(4) 2	4	12	(4) 7	(3) 3	(1) 3	13
6	(2) 1	(4) 1	(5) 5	7	13	(5) 9	(4) 5	(2) 3	17
7	(3) 2	(5) 8	(6) 9	19	14	(6) 5	(5) 3	(3) 1	9
Column totals	20	26	32	78		35	26	21	82

Basic block	Rep 1	Rep 2	Total
1	$12 = P_1$	$7 = P_8$	$19 = B_1$
2	18	17	35
3	6	13	19
4	12	6	18
5	4	13	17
6	7	17	24
7	$19 = P_7$	$9 = P_{14}$	$28 = B_7$
	78	82	160

Treatments	T_i	$B_{(i)}$	$W_i = 4T_i - 6B_{(i)} + 2G$	$Q_i = 3T_i - B_{(i)}$
1	17	71	-38	-20
2	13	61	6	-22
3	12	63	-10	-27
4	18	60	32	-6
5	34	87	-66	15
6	27	66	32	15
7	39	72	44	45
	160	480	0	0

Table 9.8-2 Computational Procedures

$$\begin{aligned}
 (1) &= G^2/knt = 609.52 & (4) &= (\Sigma B_j^2)/kn = 653.33 \\
 (2) &= \Sigma X^2 = 876 & (5) &= [\Sigma(\text{col. total})^2]/t = 634.57 \\
 (3) &= (\Sigma T_i^2)/kn = 722.00 & (6) &= (\Sigma P_m^2)/k = 718.67 \\
 & & (7) &= \Sigma(\text{rep})^2/kt = 609.90 \\
 & & (8) &= \frac{\Sigma W_i^2}{knt(t-k)(k-1)} = 29.52 \\
 & & (9) &= \frac{\Sigma Q_i^2}{nkt} = 98.19
 \end{aligned}$$

Source	SS	df	MS
Columns	(5) - (1) = 25.05	5	
Treatments (unadj)	(3) - (1) = 112.48	6	
Blocks w. rep (adj)	(6) - (4) - (7)		
	+ (1) + (8) = 94.48	12	7.87 = E_b
Intrablock error	(2) + (4) + (7) - (3)		
	- (5) - (6) - (8) = 34.47	18	1.92 = E_e
Total	(2) - (1) = 266.48	41	

$$\begin{aligned}
 \mu &= \frac{n(E_b - E_e)}{nt(k-1)E_b + (t-k)E_e} = \frac{2(5.95)}{28(7.87) + 4(1.92)} = .0522 \\
 E'_e &= E_e[1 + (t-k)\mu] = 1.92[1 + 4(.0522)] = 2.32
 \end{aligned}$$

Table 9.8-3 Adjusted Treatment Totals and Means
(Intra- and interblock data)

Treatment	$T'_i = T_i + .522W_i$	\bar{T}'_i
1	15.02	2.5
2	13.31	2.2
3	11.48	1.9
4	19.67	3.3
5	30.55	5.1
6	28.67	4.8
7	41.30	6.9
	160.00 = G	26.7 = $G/kn = G/6$

$$SS_{\text{treat(adj)}} = \frac{\Sigma(T'_i)^2}{kn} - \frac{G^2}{knt} = 730.40 - 609.52 = 120.88$$

$$MS_{\text{treat(adj)}} = SS_{\text{treat(adj)}}/(t-1) = 120.88/6 = 20.15$$

$$F = \frac{MS_{\text{treat(adj)}}}{E'_e} = \frac{20.15}{2.32} = 8.69$$

$$F_{.99}(6, 18) = 4.01$$

the sum of the two repetitions of basic block 2, $18 + 17 = 35$, will be designated B_2 .

Summary data required for the sums of squares are given in the lower part of Table 9.8-1. An entry in column T_i is the sum of the observations on treatment i . For treatment 1,

$$\begin{array}{rcl} 3 + 5 & \text{basic block 2: subjects 2 and 9} & \\ 2 + 3 & \text{basic block 3: subjects 3 and 10} & \\ 1 + 3 & \text{basic block 5: subjects 5 and 12} & \\ \hline 17 & = T_1 & \end{array}$$

An entry in column $B_{(i)}$ is the sum of the totals for basic blocks in which treatment i appears. For treatment 1,

$$\begin{array}{rcl} 35 & \text{total for } n \text{ repetitions of basic block 2} & \\ 19 & \text{total for } n \text{ repetitions of basic block 3} & \\ 17 & \text{total for } n \text{ repetitions of basic block 5} & \\ \hline 71 & = B_{(1)} & \end{array}$$

Convenient computational symbols are defined in the upper part of Table 9.8-2. For example,

$$(4) = \frac{19^2 + 35^2 + \cdots + 28^2}{6} = 653.33,$$

$$(5) = \frac{20^2 + 26^2 + \cdots + 21^2}{7} = 634.57,$$

$$(6) = \frac{12^2 + 18^2 + \cdots + 17^2 + 9^2}{3} = 718.67,$$

$$(7) = \frac{78^2 + 82^2}{21} = 609.90.$$

In each case, the divisor is the number of observations in a quantity which is squared in the numerator. For example, the quantity 78 in (7) is the sum of the 21 observations in repetition 1.

Computational formulas for the sums of squares are given in the lower part of Table 9.8-2. Procedures for obtaining the weighting factor μ and the effective error mean square E'_e are also indicated.

Adjusted treatment totals and means are given in Table 9.8-3. Both intrablock and interblock treatment information are included in the adjustment process. Procedures for making an approximate test on the over-all treatment effects are summarized. In this case the observed F ratio exceeds the critical value for a .01-level test. Hence the experimental data reject the hypothesis that there are no differences between the treatment effects. To

Table 9.8-4 Summary of Analysis of Variance
(Intrablock treatment information only)

Source	SS	df	MS	F
Columns	(5) - (1) = 25.05	5		
Blocks w. reps (unadj)	(6) - (7) = 108.75	12		
Treat (adj for blocks)	(9) = 98.19	6	16.36	8.52
Intrablock error	(2) + (7) - (5) - (6) - (9) = 34.47	18	1.92	
Total	(2) - (1) = 266.46			

$$E'_e = E_e/E = E_e/[t(k-1)/k(t-1)]$$

$$= 1.92/.778 = 2.47$$

test, for example, the difference between the mean rating for packages 5 and 7, the test statistic has the form

$$F = \frac{(T'_5 - T'_7)^2}{2knE'_e} = \frac{(30.55 - 41.30)^2}{2(6)(2.32)} = 4.15.$$

The critical value for a .01-level test is $F_{.99}(1,18) = 8.29$.

The critical value for the Tukey hsd test on all possible differences between pairs of adjusted means for a conservative .01 significance level on the entire set of comparisons is

$$q_{.99}(7,18)\sqrt{\frac{E'_e}{kn}} = 5.79\sqrt{\frac{2.32}{6}} = 3.60.$$

Two adjusted means are said to be statistically different if their difference exceeds the value 3.60. With this overstringent decision rule, package 7 differs from 1, 2, 3, and 4, but no other differences are statistically significant at a .01-level Tukey hsd test.

Table 9.8-5 Adjusted Treatment Totals and Means
(Intrablock information only)

Treatment	$T''_i = \frac{knQ_i}{nt\lambda} + kn\bar{G}$ $= \frac{6Q_i}{14} + 6(3.81)$	$\bar{T}''_i = \frac{Q_i}{t\lambda} + \bar{G}$ $= \frac{Q_i}{14} + 3.81$
1	14.29	2.4
2	13.43	2.2
3	11.35	1.9
4	20.29	3.4
5	29.29	4.9
6	29.29	4.9
7	42.15	7.0
	160.09	26.7 = $G/kn = G/6$

If only the intrablock treatment information is used in estimating treatment effects, computational procedures for sums of squares are indicated in Table 9.8-4. Computational symbols are those in Table 9.8-2. A test on the over-all difference between treatment effects can be made in the analysis-of-variance table by means of the statistic

$$F = \frac{MS_{\text{treat(adj for blocks)}}}{E_e}.$$

Adjusted treatment totals and means for this case are summarized in Table 9.8-5. A test on the difference $\tau_2 - \tau_5$, for example, has the form

$$F = \frac{(T''_2 - T''_5)^2}{2nkE''_e}.$$

In this example, the analyses with or without the use of interblock treatment information give similar results.

9.9 Partially Balanced Designs

All the designs which have been discussed so far in this chapter are balanced in the sense that each treatment is paired with every other treatment an equal number of times within a common block; that is, λ was a constant for all treatments. These balanced designs may be considered as special cases of a broader class of designs known as partially balanced incomplete block designs. In the latter some pairs occur together within a common block more often than do others. Corresponding to each treatment are two or more classes of associates.

To illustrate what is meant by a class of associates, consider the following design, in which the number of treatments $t = 10$, the block size $k = 4$, the number of blocks $b = 5$, and the number of replications $r = 2$:

Block				
<i>a</i>	1	2	3	4
<i>b</i>	1	5	6	7
<i>c</i>	2	5	8	9
<i>d</i>	3	6	8	10
<i>e</i>	4	7	9	10

Note that treatment 1 occurs in the same block with treatments 2 through 7, but treatment 1 does not occur in the same block with treatments 8, 9, or 10. Treatments 2 through 7 are said to be the first associates of treatment 1; treatments 8, 9, and 10 are said to be the second associates of treatment 1. Alternatively, the first associate class of treatment 1 is the set of treatments

2 through 7, and the second associate class of treatment 1 is the set of treatments 8, 9, and 10. A partial listing of associate classes for this design is given below:

Treatment	First associates	Second associates
1	2, 3, 4, 5, 6, 7	8, 9, 10
2	1, 3, 4, 5, 8, 9	6, 7, 10
...
9	2, 4, 5, 7, 8, 10	1, 3, 6
10	3, 4, 6, 7, 8, 9	1, 2, 5

In general if treatments i and j are first associates, then the following conditions must hold.

1. The number of treatments common to the respective first associate classes (p_{11}^1) is a constant for all i 's and j 's. For the design given above this constant is $p_{11}^1 = 3$. In the above example, the first associate classes of treatments 1 and 2 have treatments 4, 5, and 9 in common; the first associate classes of 2 and 9 have treatments 4, 5, and 8 in common.

2. The number of treatments common to the second associate classes is a constant for all i 's and j 's. In this case, the constant is $p_{22}^1 = 1$. In the above example, treatment 7 is common to the second associates of treatments 1 and 2; treatment 6 is common to second associates of 2 and 9.

3. The number of treatments common to the first associates of i and the second associates of j (or vice versa) is constant. In this case $p_{12}^1 = 2$. In the above example, the first associate class of treatment 1 and the second associate class of treatment 2 have 6 and 7 in common; the first associate class of 2 and the second associate class of 9 have 1 and 3 in common.

If treatments i and j are second associates, then the following conditions are required for partial balance:

1. The number of treatments common to the respective first associates is constant for all i 's and j 's. In the example, treatments 1 and 9, which are second associates, have four treatments in common in their respective first associate classes; similarly, treatments 2 and 10 have four treatments in common in their first associate classes. Hence, for this case, $p_{11}^2 = 4$.

2. The number of treatments common to respective second associate classes is constant. In the example, the second associate classes of 1 and 10 have no treatments in common; similarly the second associate classes of 1 and 9 have no treatments in common. In this case $p_{22}^2 = 0$.

3. The number of treatments common to the first associate class of i and the second associate class of j (or vice versa) is constant. In the example, $p_{12}^2 = 2$.

In the general case of partially balanced incomplete-block designs, each treatment may have n associate classes. Treatments in the first associate class of treatment i are paired with treatment i in λ_1 blocks; treatments in

the second associate class of i are paired with i in λ_2 blocks; . . . ; treatments in the n th associate class of i are paired with i in λ_n blocks. Balanced incomplete-block designs form that special case of partially balanced designs in which there is one associate class for each treatment. The symmetries imposed upon the structure of partially balanced designs are what permit simplified solution of the normal equations obtained in using the least-squares principle in the estimation process.

Partially balanced designs may be analyzed either with or without the recovery of interblock treatment information. Principles underlying the analysis will be found in a readable article by Bose and Shimamoto (1952). Extensive tables of partially balanced designs have been compiled by Bose et al. (1954). Since all treatments are not paired an equal number of times within a common block, in comparing pairs of treatments the proper error term depends upon the kind of association between the pair in question. One of the major disadvantages of partially balanced designs, in contrast to completely balanced designs, is the fact that differences between all pairs of treatments cannot be estimated with the same degree of precision. However, this latter disadvantage can, in part, be turned into an advantage. If greater precision is desired for some comparisons, the experimenter may assign treatment numbers to the treatments in such a way that pairs on which greater precision is desired are those which occur together in the same block most frequently.

Cases of Special Interest. The following partially balanced incomplete-block design resembles a Youden square—each of the columns defines a replication. In this design, however, each treatment has two associate classes.

Block				
<i>a</i>	1	4	2	5
<i>b</i>	2	5	3	6
<i>c</i>	3	6	1	4
<i>d</i>	4	1	5	2
<i>e</i>	5	2	6	3
<i>f</i>	6	3	4	1

A partial listing of the associate classes is as follows:

Treatment	First associates	Second associates
1	4	2, 3, 5, 6
2	5	1, 3, 4, 6
...
6	3	1, 2, 5, 6

The first associates appear together in $\lambda_1 = 4$ blocks, whereas the second associates appear together in $\lambda_2 = 2$ blocks. For example, treatments 2 and 5, which are first associates, appear together in blocks *a*, *b*, *d*, and *e*. The Youden square is that special case of this design in which all treatments have only one associate class. This plan is actually superior to a Youden square for those situations in which greater precision is required for first associates than for second associates.

In addition to the two kinds of balanced two-dimensional lattice designs that have been discussed, there are also partially balanced two-dimensional lattices. For example, there are designs in which k^2 treatments are arranged in blocks of size k with $(k + 1)/2$ replications. In the latter design, each treatment has two associate classes. Lattice-square designs may also be constructed with $(k + 1)/2$ replications. In two-dimensional lattice designs, a double lattice is one in which the number of replications is a multiple of 2; a triple lattice is one in which the number of replications is a multiple of 3. Consider the following triple lattice (since $t = 9 = 3^2 = k^2$, the following design may be classified as two-dimensional):

Replication 1				Replication 2				Replication 3			
Block				Block				Block			
<i>a</i>	1	2	3	<i>d</i>	2	6	7	<i>g</i>	1	4	7
<i>b</i>	4	5	6	<i>e</i>	1	5	9	<i>h</i>	3	6	9
<i>c</i>	7	8	9	<i>f</i>	3	4	8	<i>i</i>	2	5	8

In this design, each treatment has two associate classes. For the first associates $\lambda_1 = 1$; for second associates $\lambda_2 = 0$. Each treatment has only a single second associate. Hence this triple lattice is close to being balanced.

In three-dimensional lattice designs k^3 treatments are arranged in blocks of size k . This type of design may be related to a $k \times k \times k$ factorial experiment arranged in blocks of size k . The principles underlying the analysis of variance for this case are essentially those of the corresponding factorial experiment. A complete replication requires k^2 blocks. Partially balanced designs of this kind require a minimum of three, or some multiple of three, replications. The computational procedures for a three-dimensional triple lattice for $k^3 = 27$ treatments in blocks of size $k = 3$ with $r = 3$ replications are given in Federer (1955, pp. 349–355).

Rectangular lattices form a broad category under the partially balanced incomplete-block designs. These designs may be considered as two-dimensional lattice designs having one restriction, but the number of treatments is kp . Of special interest are the so-called rectangular lattices having $k(k + 1)$ treatments in blocks of size k . The following design represents a triple lattice (i.e., the number of replications is a multiple of 3) having

$k(k + 1) = 12$ treatments arranged in blocks of size $k = 3$:

Replication 1				Replication 2				Replication 3			
Block				Block				Block			
<i>a</i>	1	2	3	<i>f</i>	1	8	11	<i>j</i>	1	5	7
<i>b</i>	4	5	6	<i>g</i>	2	5	12	<i>k</i>	2	9	10
<i>c</i>	7	8	9	<i>h</i>	3	6	9	<i>l</i>	3	4	11
<i>d</i>	10	11	12	<i>i</i>	4	7	10	<i>m</i>	6	8	12

In the above design, treatments which are first associates appear together in $\lambda_1 = 1$ block; treatments which are second associates appear together in $\lambda_2 = 0$ block. The general procedures for handling partially balanced incomplete-block designs may be used for rectangular lattices. However, the general computational formulas can be simplified for these special designs.

Designs for $k = 2$. Of special interest in the behavioral sciences are designs having block size $k = 2$. A block may consist of a pair of twins, a husband and wife, the right and left hand of an individual, etc. A balanced design for the case $k = 2$ requires a minimum of $t(t - 1)/2$ blocks. For example, a balanced design for the case in which $t = 10$ requires a minimum of 45 blocks. If the number of treatments is at all large, the only feasible design may be a partially balanced design rather than a balanced design.

A special class of partially balanced designs for the case in which $k = 2$ has been developed by Zoellner and Kempthorne (1954). These designs are known as circulant designs. Although the number of associate classes for these designs increases as the number of treatments increases, the computational work for circulant designs is relatively simple when the design constants associated with the reduced normal equations are known. The latter constants have been tabulated for cases in which the design corresponds to specified plans.

A plan for a design in which $t = 7$ and $r = 4$ is specified by the following set of numbers:

$$* \quad 1 \quad 1 \quad 0 \quad 0 \quad 1 \quad 1$$

To determine the treatments which occur in the same block with treatment 1, for example, one adds to the digit 1 the numbers 1, 2, —, —, 5, 6. The latter set of numbers corresponds to the pattern (1 1 0 0 1 1). Thus the treatments which appear in the same block with treatment 1 are treatments 2, 3, 6, and 7. To find the treatments appearing in the same block with treatment 5, one adds to the digit 5 the same set of numbers, subtracting $t = 7$ if the resulting sum exceeds 7. Thus the treatments paired with treatment 5 are 6, 7, 3, and 4. In general, the treatments paired with treatment j are $j + 1, j + 2, j + 5$, and $j + 6, t = 7$ being subtracted when the sum exceeds 7.

A plan for a design in which $t = 8$ and $r = 6$ may be specified by the following set of numbers:

$$* \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 1$$

The digits to be added to a treatment number are

$$1 \quad 2 \quad 3 \quad - \quad 5 \quad 6 \quad 7$$

Thus the treatments which appear in the same block with treatment j are $j + 1, j + 2, j + 3, j + 5, j + 6, j + 7$.

If a sum exceed $t = 8$, then the corresponding treatment is defined to be the sum minus 8. The design constants associated with designs constructed in accordance with these plans have been tabulated by Zoellner and Kempthorne (1954, p. 179). Some of the plans in this series for which the design

Table 9.9-1 Plans for Circulant Designs for $k = 2$

t	r	Plan
6	4	* 1 1 0 1 1
7	4	* 1 1 0 0 1 1
8	6	* 1 1 1 0 1 1 1
8	5	* 1 1 0 1 0 1 1
8	4	* 1 0 1 0 1 0 1
8	3	* 1 0 0 1 0 0 1
9	6	* 1 1 0 1 1 0 1 1
9	4	* 1 0 1 0 0 1 0 1
10	5	* 1 0 1 0 1 0 1 0 1
10	4	* 1 0 0 1 0 1 0 0 1

constants have been tabulated are given in Table 9.9-1. The structure of the blocks for the design in which $t = 6$ and $r = 4$, having the plan (* 1 1 0 1 1), is as follows:

(1,2) (1,3) (1,5) (1,6) (2,3) (2,4) (2,6) (3,4) (3,5) (4,5) (4,6) (5,6)

The efficiency factor for the designs given in Table 9.9-1 is in the range from .40 to .60. A numerical example of the computational procedures for these designs is given by Zoellner and Kempthorne.

9.10 Numerical Example of Partially Balanced Design

Although most of the examples that have been given in this chapter have been cast in terms of experiments having repeated measures, it should be noted that a block may represent any source of variation that an experimenter may wish to control—a source of experimental material, a time of day, an experimenter. The alternative to arranging an experiment in blocks is complete randomization. This alternative will serve, in the long run, to avoid confounding of treatment effects. However, the variation due to the potential blocks becomes part of the experimental error.

When the experimental material is arranged in blocks, variation due to the latter source is not part of the experimental error.

The numerical example to be considered in this section is for the case in which the number of treatments $t = 10$, the block size $k = 4$, and the number of replications $r = 4$. With tables of this design are the design constants $c_1 = \frac{4}{15}$, $c_2 = \frac{8}{15}$, $\lambda_1 = 1$, and $\lambda_2 = 2$. The actual structure of the design as well as the criterion scores are given in Table 9.10-1.

Table 9.10-1 Numerical Example of Partially Balanced Incomplete-block Design

($t = 10$, $k = 4$, $r = 4$, $c_1 = \frac{4}{15}$, $c_2 = \frac{8}{15}$, $\lambda_1 = 1$, $\lambda_2 = 2$)

Block					B_j
1	(9)	(3)	(7)	(5)	59
	18	13	16	12	
2	(7)	(2)	(6)	(10)	85
	21	15	22	27	
3	(3)	(9)	(1)	(6)	55
	13	18	10	14	
4	(6)	(4)	(5)	(8)	61
	18	15	17	11	
5	(8)	(1)	(10)	(9)	46
	7	6	17	16	
6	(1)	(8)	(4)	(7)	34
	5	12	8	9	
7	(4)	(6)	(9)	(2)	60
	14	15	19	12	
8	(2)	(7)	(8)	(3)	76
	15	19	24	18	
9	(5)	(10)	(2)	(1)	36
	7	15	7	7	
10	(10)	(5)	(3)	(4)	78
	26	18	17	17	
					590 = G

This design has two associate classes. For example, first associates of treatment 1 are 2, 3, 4, 5, 6, and 7; treatment 1 appears with each of these treatments in $\lambda_1 = 1$ block. The second associates of treatment 1 are 8, 9, and 10; treatment 1 appears with each of these treatments in $\lambda_2 = 2$ blocks.

The columns of this design form complete replications, but the replications aspect will be disregarded in the analysis that follows—assume that the treatments within each block are arranged in a randomized sequence. Summary data are given in Table 9.10-2. An entry in column T_i is the sum of the $r = 4$ observations on treatment i . An entry in column $B_{(i)}$ is the sum of the totals for blocks in which treatment i appears. An entry in column

Table 9.10-1a Associate Classes for Design in Table 9.10-1

Treatment	First associates	Second associates
1	2, 3, 4, 5, 6, 7	8, 9, 10
2	1, 3, 4, 5, 8, 9	6, 7, 10
3	1, 2, 4, 6, 8, 10	5, 7, 9
4	1, 2, 3, 7, 9, 10	5, 6, 8
5	1, 2, 6, 7, 8, 9	3, 4, 10
6	1, 3, 5, 7, 8, 10	2, 4, 9
7	1, 4, 5, 6, 9, 10	2, 3, 8
8	2, 3, 5, 6, 9, 10	1, 4, 7
9	2, 4, 5, 7, 8, 10	1, 3, 6
10	3, 4, 6, 7, 8, 9	1, 2, 5

$P_{11}^1 = 3$	$P_{12}^1 = 2$	$P_{11}^2 = 4$	$P_{12}^2 = 2$
$P_{21}^1 = 2$	$P_{22}^1 = 1$	$P_{21}^2 = 2$	$P_{22}^2 = 0$

Table 9.10-2 Summary Data

Treatment	T_i	$B_{(i)}$	Q_i	2d assoc.	$\sum_{(i)} Q_i$	(a)	(b)	(c)
1	28	171	-59	8, 9, 10	158	-99	-668	11.0
2	49	257	-61	6, 7, 10	116	-55	-738	10.7
3	61	268	-24	5, 7, 9	52	-28	-284	13.2
4	54	233	-17	5, 6, 8	-4	21	-242	13.4
5	54	234	-18	3, 4, 10	54	-36	-198	13.7
6	69	261	15	2, 4, 9	-14	-1	196	15.8
7	65	254	6	2, 3, 8	-86	80	-2	14.7
8	54	217	-1	1, 4, 7	-70	71	-84	14.3
9	71	220	64	1, 3, 6	-68	4	828	19.3
10	85	245	95	1, 2, 5	-138	43	1192	21.4
	590	2360	0		0	0	0	147.5

$$G = 14.75$$

$$(a) = -[Q_i + \sum_{(i)} Q_i]$$

$$(b) = 13Q_i - (a) = 14Q_i + \sum_{(i)} Q_i - 180r_i$$

$$(c) = [(b)/180] + 14.75 = T_i$$

Q_i has the general form

$$Q_i = kT_i - B_{(i)}$$

The symbol $\sum_{(i)} Q_i$ will be used to designate the sum of those Q_i 's which correspond to first associates of treatment i . For example,

$$\sum_{(1)} Q_i = Q_2 + Q_3 + Q_4 + Q_5 + Q_6 + Q_7 = -99,$$

$$\sum_{(2)} Q_i = Q_1 + Q_3 + Q_4 + Q_5 + Q_6 + Q_8 = -55.$$

These sums appear in column *a* in the table. There is a simpler method for obtaining these sums. Suppose that the symbol $\sum_{(i)} Q_i$ represents the sum of the Q_i 's corresponding to the treatments which are second associates of treatment *i*. It is readily shown that

$$\sum_{(i)} Q_i + \sum_{(i)} Q_i + Q_i = 0.$$

For example, for treatment 1,

$$\sum_{(1)} Q_1 + \sum_{(1)} Q_1 + Q_1 = -99 + 158 + (-59) = 0.$$

Hence,

$$\sum_{(i)} Q_i = -(\sum_{(i)} Q_i + Q_i).$$

In this design each treatment has three second associates but six first associates. Thus the computation of $\sum_{(i)} Q_i$ involves less work than does the computation of $\sum_{(i)} Q_i$. The former are given in the column $\sum_{(i)} Q_i$ of the table, the latter in column *a*. It will be found that entries in just one of these two columns need to be obtained.

The intrablock estimate of the treatment effect is in general

$$rk(k-1)t_i = (k - c_k)Q_i + (c_k - c_k)\sum_{(i)} Q_i$$

For the design in Table 9.10-1,

$$48t_i = (4 - \frac{1}{3})Q_i + (\frac{1}{3} - \frac{1}{3})\sum_{(i)} Q_i$$

which becomes, upon clearing fractions,

$$720t_i = 52Q_i - 4\sum_{(i)} Q_i$$

Both sides may now be divided by 4 to give

$$180t_i = 13Q_i - \sum_{(i)} Q_i$$

Using the relation between $\sum_{(i)} Q_i$ and $\sum_{(i)} Q_i$,

$$\begin{aligned} 180t_i &= 13Q_i + Q_i + \sum_{(i)} Q_i \\ &= 14Q_i + \sum_{(i)} Q_i \end{aligned}$$

Entries in column *b* of Table 9.10-2 are obtained most readily from this last expression.

Estimates of adjusted treatment means, by using intrablock information, have the form

$$\bar{T}_i^* = t_i + G.$$

Each entry in column *b* estimates $180\bar{T}_i^*$. Hence

$$\bar{T}_i^* = \frac{\text{entry in col } b}{180} + G.$$

Computational formulas for the sums of squares are given in Table 9.10-3. Since each entry in column b of Table 9.10-2 is $180t_i$, symbol (5) is computed as follows:

$$\begin{aligned}(5) &= \frac{1}{180k} \sum (180t_i)Q_i \\ &= \frac{1}{20} [(-668)(-59) + (-738)(-61) + \cdots + (1192)(95)] \\ &= 372.46.\end{aligned}$$

Table 9.10-3 Computational Formulas for Sums of Squares

$(1) = G^2/rt = 8702.50$
 $(2) = \sum X^2 = 9826$

$(3) = (\sum T_i^2)/r = 9226.50$
 $(4) = (\sum B_j^2)/k = 9370.00$

$(5) = \frac{1}{k} \sum t_i Q_i = 372.46$

Source	SS	df	MS	F
Blocks (unadj)	(4) - (1) = 667.50	9	115.55	10.40*
Treatments (adj for blocks)	(5) = 372.46	9	41.38	
Intrablock error	(2) - (4) - (5) = 83.54	21	3.98 = E_e	
Total	(2) - (1) = 1123.50	39		

Treatments (unadj) (3) - (1) = 524.00

* $F_{.95}(9,21) = 2.37$

$$E_1'' = E_e \left(\frac{k - c_1}{k - 1} \right) = 3.98 \left(\frac{4 - \frac{4}{5}}{3} \right) = 4.95$$

$$E_2'' = E_e \left(\frac{k - c_2}{k - 1} \right) = 3.98 \left(\frac{4 - \frac{8}{5}}{3} \right) = 4.60$$

$$\bar{E}'' = \frac{n_1 E_1'' + n_2 E_2''}{n_1 + n_2} = \frac{6(4.95) + 3(4.60)}{9} = 4.83$$

The analysis of variance is also summarized in this table. A test on the over-all hypothesis that $\sigma_e^2 = 0$ may be made from the data in the analysis-of-variance table—the latter utilizes only the intrablock treatment information.

Effective error terms for comparisons between treatments which are first or second associates are also indicated. When these two errors do not differ appreciably, an average error may be computed as

$$\bar{E}_e'' = \frac{n_1 E_1'' + n_2 E_2''}{n_1 + n_2},$$

where n_1 and n_2 are, respectively, the number of first and second associates for any treatment. Comparisons between two treatments which are first associates have the following form:

$$F = \frac{(T_1'' - T_2'')^2}{2E_1''/r}.$$

Comparisons between two treatments which are second associates have the following form:

$$F = \frac{(T_1'' - T_2'')^2}{2E_2''/r}.$$

It should be noted that, when $\lambda_1 < \lambda_2$, E_1'' will be larger than E_2'' .

When the interblock variance is large relative to the intrablock variance (in this case the interblock variance is quite large relative to E_2), the interblock treatment information may generally be disregarded. Computational details for the recovery of interblock treatment information are given in Cochran and Cox (1957, pp. 460-463).

9.11 Linked Paired-comparison Designs

Designs having block size equal to 2 are of particular interest in paired-comparison work. Bose (1956) has described a special class of designs, which he calls linked paired-comparison designs. These designs are related to balanced incomplete-block designs. The notation used by Bose is as follows:

n = number of objects to be compared.

v = number of judges to be used.

r = number of pairs of objects to be compared by each judge.

$b = n(n-1)/2$ = number of pairs.

The following conditions are required.

1. For the r pairs of objects compared by each of the judges, each object must appear α times.

2. Each pair of objects is judged by k judges, where $k > 1$.

3. For any two judges, there are exactly λ pairs of objects which are common to the r pairs assigned to each judge.

The following design is an example of a linked paired-comparison design in which $n = 4$, $v = 3$, $b = 6$, $r = 4$, $k = 2$, $\lambda = 2$, and $\alpha = 2$:

Judge	Pairs assigned to judge
<i>a</i>	(1,4), (1,3), (2,4), (2,3)
<i>b</i>	(1,3), (2,4), (1,2), (3,4)
<i>c</i>	(1,4), (1,2), (2,3), (3,4)

Among the $r = 4$ pairs of objects assigned to a judge, each object appears $\alpha = 2$ times. Any two judges have $\lambda = 2$ pairs in common. For example, judge *a* and judge *b* have the pairs (1,3) and (2,4) in common.

There is a correspondence between linked block designs and balanced incomplete-block designs. Bose indicates a method for obtaining the former from the latter. In this correspondence, a judge has the role of a treatment, and a set of pairs of treatments has the role of a block. To illustrate this correspondence, suppose that n is an even number and that the symbols with asterisks represent parameters of a balanced incomplete-block design; then the parameters of a linked paired-comparison design are

related to those of a balanced incomplete-block design (marked with an asterisk) as follows:

$$v = v^*, \quad r = \frac{nr^*}{2}, \quad k = k^*, \quad \lambda = \frac{n\lambda^*}{2}, \quad \alpha = r^*.$$

(In this context v^* represents the number of treatments.) Thus, if

$$n = 8, \quad v = 7, \quad r = 12, \quad k = 3, \quad \lambda = 4, \quad \text{and } \alpha = 3,$$

for a linked paired-comparison design, then

$$v^* = 7, \quad r^* = 3, \quad k^* = 3, \quad \lambda^* = 1$$

for the corresponding balanced incomplete-block design. In the latter design,

$$b^* = \frac{v^*r^*}{k^*} = \frac{7(3)}{3} = 7,$$

To illustrate the method of obtaining the linked paired-comparison design from the balanced incomplete-block design described in the last paragraph, suppose that the latter design is represented as follows:

<i>Blocks</i>	<i>Treatments</i>		
I	<i>a</i>	<i>b</i>	<i>d</i>
II	<i>b</i>	<i>c</i>	<i>e</i>
III	<i>c</i>	<i>d</i>	<i>f</i>
IV	<i>d</i>	<i>e</i>	<i>g</i>
V	<i>e</i>	<i>f</i>	<i>a</i>
VI	<i>f</i>	<i>g</i>	<i>b</i>
VII	<i>g</i>	<i>a</i>	<i>c</i>

If $n = 8$, there are $8(7)/2 = 28$ pairs of objects. These pairs may be arranged in seven sets of four pairs each in a way that no object appears twice within any set. The sets are as follows (note the cyclic arrangement of the columns):

<i>Set</i>	<i>Pairs</i>
I	(4,8), (1,7), (2,6), (3,5)
II	(5,8), (2,1), (3,7), (4,6)
III	(6,8), (3,2), (4,1), (5,7)
IV	(7,8), (4,3), (5,2), (6,1)
V	(1,8), (5,4), (6,3), (7,2)
VI	(2,8), (6,5), (7,4), (1,3)
VII	(3,8), (7,6), (1,5), (2,4)

Suppose that the treatments in the balanced incomplete-block designs are considered to be judges; suppose that the judges are assigned the sets of pairs corresponding to the block in which the corresponding treatment is located. That is, since treatment a appears in blocks I, V, and VII, the corresponding sets of pairs are assigned to judge a . Similarly, since treatment b appears in blocks I, II, and VI, the corresponding sets of pairs are assigned to judge b .

In this way each judge is assigned three sets of four pairs each. Hence each judge makes $r = 12$ comparisons. In a complete block design, each judge would make 28 comparisons.

Bose has not suggested special methods for analyzing data obtained from a linked block paired-comparison design. There are several general methods for analyzing paired-comparison data. A review of these methods as well as an extensive bibliography will be found in Jackson and Fleckenstein (1957). In particular, the method suggested by Scheffé could be adapted for handling the analysis of data obtained from linked block paired-comparison designs. These designs should prove to be particularly useful in psychometric scaling.

CHAPTER 10

Latin Squares and Related Designs

10.1 Definition of Latin Square

A Latin square, as used in experimental design, is a balanced two-way classification scheme. Consider the following 3×3 arrangement:

<i>a</i>	<i>b</i>	<i>c</i>
<i>b</i>	<i>c</i>	<i>a</i>
<i>c</i>	<i>a</i>	<i>b</i>

In this arrangement, each letter occurs just once in each row and just once in each column. The following arrangement also exhibits this kind of balance:

<i>b</i>	<i>c</i>	<i>a</i>
<i>a</i>	<i>b</i>	<i>c</i>
<i>c</i>	<i>a</i>	<i>b</i>

The latter arrangement was obtained from the first by interchanging the first and second rows.

Use of this type of balance may be incorporated into a variety of designs. For example, consider an experiment involving the administration of three treatments to each of three subjects. The order in which subjects receive the treatments may be either completely randomized or randomized under the restriction of balance required for the Latin square. If the treatments are designated a_1 , a_2 , and a_3 , a balanced arrangement with respect to order of administration is obtained by employing the following plan:

	Order 1	Order 2	Order 3
Subject 1	a_1	a_2	a_3
Subject 2	a_3	a_1	a_2
Subject 3	a_2	a_3	a_1

If this plan is followed, each treatment is administered first once, second once, and third once. If there is a systematic additive effect associated with the order of administration, this effect can be evaluated. Had order of

administration been randomized independently for each subject, this balance would not in general be achieved and evaluation of the order effect could not be made readily. (In the design outlined above, instead of individuals, groups of subjects may be assigned to the rows of the Latin square.)

Two Latin squares are orthogonal if, when they are combined, the same pair of symbols occurs no more than once in the composite square. For example, consider the following 3×3 Latin squares:

(1)			(2)			(3)		
a_1	a_2	a_3	b_2	b_3	b_1	c_1	c_2	c_3
a_2	a_3	a_1	b_3	b_1	b_2	c_3	c_1	c_2
a_3	a_1	a_2	b_1	b_2	b_3	c_2	c_3	c_1

Combining squares (1) and (2) yields the composite square

a_1b_2	a_2b_3	a_3b_1
a_2b_3	a_3b_1	a_1b_2
a_3b_1	a_1b_2	a_2b_3

In this composite the treatment combination a_1b_2 occurs more than once; hence squares (1) and (2) are not orthogonal.

Combining squares (1) and (3) yields the following composite:

a_1c_1	a_2c_2	a_3c_3
a_2c_3	a_3c_1	a_1c_2
a_3c_2	a_1c_3	a_2c_1

In this composite no treatment combination is repeated. There are nine possible treatment combinations that may be formed from three levels of factor A and three levels of factor C . Each of these possibilities appears in the composite. Hence squares (1) and (3) are orthogonal.

Extensive tables of sets of orthogonal Latin squares are given in Fisher and Yates (1953) and Cochran and Cox (1957). The composite square obtained by combining two orthogonal Latin squares is called a Greco-Latin square. The 3×3 Greco-Latin square obtained by combining squares (1) and (3) above may be represented schematically as follows:

11	22	33
23	31	12
32	13	21

This representation uses only the subscripts that appear with the a 's and c 's. Interchanging any two rows of a Greco-Latin square will still yield a Greco-Latin square. For example, interchanging the first and third rows of the above square yields the following Greco-Latin square:

32	13	21
23	31	12
11	22	33

Any two columns of a Greco-Latin square may also be interchanged without affecting the required balance.

It is not always possible to find a Latin square orthogonal to a given Latin square. For example, no orthogonal squares exist for 6×6 squares. Also, no orthogonal squares exist for 10×10 squares. If the dimension of a Latin square is capable of being expressed in the form (prime number)^{*n*}, where *n* is any integer, then orthogonal squares exist. For example, orthogonal Latin squares exist for squares of the following dimensions:

$$3 = 3^1, \quad 4 = 2^2, \quad 5 = 5^1, \quad 8 = 2^3, \quad 9 = 3^2.$$

However, neither 6 nor 10 may be expressed in this form. In addition, there are cases in which orthogonal squares exist when the dimension of the square is not of the form (prime number)^{*n*}—particularly when the dimension is divisible by 4. For example, orthogonal squares may be constructed for the 12×12 square.

For a $p \times p$ square which has orthogonal squares, there are at most $p - 1$ squares in an orthogonal set. The latter is a collection of $p \times p$ squares which are mutually orthogonal to each other; i.e., each square is orthogonal to all others in the set. Three or more orthogonal squares may be combined to yield hyper-Latin squares. Such squares are readily obtained from tables of complete sets of orthogonal squares.

10.2 Enumeration of Latin Squares

The standard form of a Latin square is, by definition, that square obtained by rearranging the rows and columns until the letters in the first row and the letters in the first column are in alphabetical order. For example, the square

$$(1) \quad \begin{array}{ccc} b & a & c \\ c & b & a \\ a & c & b \end{array}$$

has the standard form

$$(2) \quad \begin{array}{ccc} a & b & c \\ b & c & a \\ c & a & b \end{array}$$

The standard form (2) is obtained from (1) by interchanging columns 1 and 2. All 3×3 Latin squares may be reduced to this standard form. From this standard form $(3!) \cdot (2!) - 1 = (3 \cdot 2 \cdot 1)(2 \cdot 1) - 1 = 11$ different nonstandard 3×3 Latin squares may be constructed. Hence, including the standard form, there are 12 different 3×3 Latin squares.

For 4×4 Latin squares there are 4 different standard forms. One of these 4 is given in (3).

$$(3) \quad \begin{array}{cccc} a & b & c & d \\ b & a & d & c \\ c & d & b & a \\ d & c & a & b \end{array}$$

A second standard form is given by (4).

(4)

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>b</i>	<i>c</i>	<i>d</i>	<i>a</i>
<i>c</i>	<i>d</i>	<i>a</i>	<i>b</i>
<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>

From each of these standard forms $(4!)(3!) - 1 = 143$ different nonstandard 4×4 Latin squares may be constructed by the process of interchanging rows and columns. Nonstandard squares constructed from standard form (3) will be different from nonstandard squares constructed from standard form (4). Thus (3) potentially represents 144 Latin squares (143 nonstandard squares plus 1 standard square). Standard form (4) also represents potentially 144 different 4×4 Latin squares. Since there are 4 different standard forms, the potential number of different 4×4 Latin squares is $4(144) = 576$.

The number of possible standard forms of a Latin square increases quite rapidly as the dimension of the square increases. For example, a 6×6 Latin square has 9408 standard forms. Each of these forms represents potentially $(6!)(5!) = 86,400$ different squares. The total number of different 6×6 Latin squares that may be constructed is

$$(9408)(86,400) = 812,851,200.$$

In most tables, only standard forms of Latin squares are given. To obtain a Latin square for use in an experimental design, one of the standard squares of suitable dimension should be selected at random. The rows and columns of the selected square are then randomized independently. The levels of the factorial effects are then assigned at random to the rows, columns, and Latin letters of the square, respectively. To illustrate the procedure for randomizing the rows and columns of a 4×4 Latin square, suppose that square (3) is obtained from a table of Latin squares. Two random sequences of digits 1 through 4 are then obtained from tables of random sequences. Suppose that the sequences obtained are (2,4,1,3) and (3,4,1,2). The columns of square (3) are now rearranged in accordance with the first sequence. That is, column 2 is moved into the first position, column 4 into the second position, column 1 into the third position, and column 3 into the fourth position. The resulting square is

<i>b</i>	<i>d</i>	<i>a</i>	<i>c</i>
<i>a</i>	<i>c</i>	<i>b</i>	<i>d</i>
<i>d</i>	<i>a</i>	<i>c</i>	<i>b</i>
<i>c</i>	<i>b</i>	<i>d</i>	<i>a</i>

The rows of the resulting squares are now rearranged in accordance with the second random sequence. That is, row 3 is moved to the first position, row 4 moved to the second position, row 1 moved to the third position, and row

2 moved to the fourth position. After these moves have been made, the resulting square is

<i>d</i>	<i>a</i>	<i>c</i>	<i>b</i>
<i>c</i>	<i>b</i>	<i>d</i>	<i>a</i>
<i>b</i>	<i>d</i>	<i>a</i>	<i>c</i>
<i>a</i>	<i>c</i>	<i>b</i>	<i>d</i>

This last square represents a random rearrangement of the rows and columns of the original 4×4 standard form. This last square may be considered as a random choice from among the 144 squares represented by the standard form (3). Since (3) was chosen at random from the standard forms, the square that has just been constructed may be considered a random choice from the 576 possible different 4×4 Latin squares.

Some additional definitions of types of Latin squares will be helpful in using tables. The *conjugate* of a Latin square is obtained by interchanging the rows and columns. For example, the following squares are conjugates:

<i>a</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>b</i>
<i>c</i>	<i>a</i>	<i>b</i>	<i>b</i>	<i>a</i>	<i>c</i>
<i>b</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>b</i>	<i>a</i>

The square on the right is obtained from the square on the left by writing the columns as rows. That is, the first column of the square on the left is the first row of the square on the right, the second column of the square on the left is the second row of the square on the right, etc. Conversely, the square on the left may be obtained from the square on the right by writing the rows of the latter as columns.

A one-step *cyclic permutation* of a sequence of letters is one which moves the first letter in the sequence to the extreme right, simultaneously moving all other letters in one position to the left. For example, given the sequence *abcd*, a one-step cyclic permutation yields *bcda*; a second cyclic permutation yields *cdab*; a third cyclic permutation yields *dabc*; and a fourth cyclic permutation yields *abcd*—the latter is the starting sequence. Given a sequence of p letters, a $p \times p$ Latin square may be constructed by $p - 1$ one-step cyclic permutations of these letters. In the case of the sequence *abcd*, the Latin square formed by cyclic permutations is

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>b</i>	<i>c</i>	<i>d</i>	<i>a</i>
<i>c</i>	<i>d</i>	<i>a</i>	<i>b</i>
<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>

A *balanced set* of Latin squares is a collection in which each letter appears in each possible position once and only once. For a $p \times p$ square, there will be p squares in a balanced set. Given any $p \times p$ square, a balanced set may be constructed by cyclic permutations of the columns.

For example, the squares given below form a balanced set:

I			II			III		
<i>a</i>	<i>b</i>	<i>c</i>	<i>b</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>a</i>	<i>b</i>
<i>c</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>c</i>	<i>b</i>	<i>c</i>	<i>a</i>
<i>b</i>	<i>c</i>	<i>a</i>	<i>c</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>c</i>

Square II is obtained from square I by moving the first column of square I to the extreme right. Square III is obtained from square II by moving the first column of square II to the extreme right. The resulting set is balanced in the sense that each of the nine positions within the squares contains any given letter once and only once.

One square may be obtained from another by substitution of one letter for another. For example, given the following square:

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>b</i>	<i>a</i>	<i>d</i>	<i>c</i>
<i>c</i>	<i>d</i>	<i>b</i>	<i>a</i>
<i>d</i>	<i>c</i>	<i>a</i>	<i>b</i>

Suppose that each of the *a*'s is replaced by a *b*, each of the *b*'s replaced by a *c*, each of the *c*'s by a *d*, and each of the *d*'s replaced by an *a*. This kind of replacement procedure is called a *one-to-one transformation*. The square resulting from this transformation is given below:

<i>b</i>	<i>c</i>	<i>a</i>	<i>d</i>
<i>c</i>	<i>b</i>	<i>d</i>	<i>a</i>
<i>a</i>	<i>d</i>	<i>c</i>	<i>b</i>
<i>d</i>	<i>a</i>	<i>b</i>	<i>c</i>

This last square has the standard form given below:

<i>a</i>	<i>b</i>	<i>c</i>	<i>d</i>
<i>b</i>	<i>d</i>	<i>a</i>	<i>c</i>
<i>c</i>	<i>a</i>	<i>d</i>	<i>b</i>
<i>d</i>	<i>c</i>	<i>b</i>	<i>a</i>

Since this latter square has a different standard form from that of the original square, the one-to-one transformation yields a square which is different from any square which can be constructed by permutation of the rows or columns of the original square.

10.3 Structural Relation between Latin Squares and Three-factor Factorial Experiments

There is an interesting structural relation between a $p \times p \times p$ factorial experiment and a $p \times p$ Latin square. This relation will be illustrated by means of a $3 \times 3 \times 3$ factorial experiment. It will be shown that the latter may be partitioned into a balanced set of 3×3 Latin squares. In the following representation of a $3 \times 3 \times 3$ factorial experiment, those treatment combinations marked *X* form one Latin square, those marked

Y form a second, and those marked Z form a third. The set of three squares formed in this manner constitutes a balanced set:

	c_1			c_2			c_3		
	b_1	b_2	b_3	b_1	b_2	b_3	b_1	b_2	b_3
a_1	X	Y	Z	Z	X	Y	Y	Z	X
a_2	Z	X	Y	Y	Z	X	X	Y	Z
a_3	Y	Z	X	X	Y	Z	Z	X	Y

The treatment combinations marked X may be grouped as follows:

abc_{111}	abc_{122}	abc_{133}
abc_{221}	abc_{232}	abc_{213}
abc_{331}	abc_{312}	abc_{323}

This set of treatment combinations may also be represented in the following schematic form:

	(X)		
	b_1	b_2	b_3
a_1	c_1	c_2	c_3
a_2	c_3	c_1	c_2
a_3	c_2	c_3	c_1

In this latter arrangement the c 's form a Latin square.

The treatment combinations marked Y and Z in the factorial experiment may be represented by the following Latin squares:

(Y)				(Z)			
	b_1	b_2	b_3		b_1	b_2	b_3
a_1	c_3	c_1	c_2	a_1	c_2	c_3	c_1
a_2	c_2	c_3	c_1	a_2	c_1	c_2	c_3
a_3	c_1	c_2	c_3	a_3	c_3	c_1	c_2

In a complete factorial experiment there is balance in the sense that treatment a_i occurs in combination with each b_j and each c_k as well as in combination with all possible pairs bc_{jk} . In a Latin square there is only partial balance; each a_i occurs in combination with each b_j and each c_k , but each a_i does not occur in combination with all possible pairs bc_{jk} . For example, in square (Y) only the following combinations of bc occur with level a_1 :

$$bc_{13} \quad bc_{21} \quad bc_{32}$$

The following combinations of bc do not occur with level a_1 :

$$bc_{11} \quad bc_{12} \quad bc_{22} \quad bc_{23} \quad bc_{31} \quad bc_{33}$$

A Latin square is balanced with respect to main effects but only partially balanced with respect to two-factor interactions.

It is of interest to note the parameters estimated by the sum of all observations at level a_1 in Latin square (Y). Assume for the moment that the cell means have the following form:

$$\begin{aligned}\overline{ABC}_{113} &= \mu + \alpha_1 + \beta_1 + \gamma_3 + \varepsilon_{113} + \alpha\beta_{11} + \alpha\gamma_{13} + \beta\gamma_{13} \\ \overline{ABC}_{121} &= \mu + \alpha_1 + \beta_2 + \gamma_1 + \varepsilon_{121} + \alpha\beta_{12} + \alpha\gamma_{11} + \beta\gamma_{21} \\ \overline{ABC}_{132} &= \mu + \alpha_1 + \beta_3 + \gamma_2 + \varepsilon_{132} + \alpha\beta_{13} + \alpha\gamma_{12} + \beta\gamma_{32} \\ \bar{A}_1 &= \mu + \alpha_1 + 0 + 0 + \bar{\varepsilon}_1 + 0 + 0 + \bar{\beta}\gamma_{\alpha_1}\end{aligned}$$

Assuming further that factors A , B , and C are fixed, it follows that

$$\sum_j \beta_j = 0, \quad \sum_k \gamma_k = 0, \quad \sum_j \alpha\beta_{1j} = 0, \quad \sum_k \alpha\gamma_{1k} = 0.$$

However, $\beta\gamma_{13} + \beta\gamma_{21} + \beta\gamma_{32} \neq 0$.

Thus, in addition to sources of variation due to α and $\bar{\varepsilon}$, \bar{A} includes a source of variation due to $\beta\gamma$. Unless $\sigma_{\beta\gamma}^2 = 0$, the main effect due to factor A will be confounded with the BC interaction.

In general, when a Latin square is considered as a fractional replication of a three-factor experiment, unless $\sigma_{\beta\gamma}^2 = 0$, the main effect due to factor A will be confounded with the BC interaction. Similarly, it may be shown that unless $\sigma_{\alpha\gamma}^2 = 0$, the main effect due to factor B will be confounded with the AC interaction; unless $\sigma_{\alpha\beta}^2 = 0$, the main effect due to factor C will be confounded with the AB interaction. The variance due to the three-factor interaction, $\sigma_{\alpha\beta\gamma}^2$, is assumed to be zero throughout.

10.4 Uses of Latin Squares

In agricultural experimentation, the cells of a $p \times p$ Latin square define p^2 experimental units. In this case, the selection of a Latin square defines the manner in which the treatments are assigned to the experimental units. That is, the procedure followed in selecting the Latin square specifies the randomization procedure necessary for the validity of statistical tests made in subsequent analyses. In its classic agricultural setting, the Latin-square design represents a single-factor experiment with restricted randomization with respect to row and column effects associated with the experimental units. In this design it is assumed that treatment effects do not interact with the row and column effects. Fisher (1951) and Wilk and Kempthorne (1957) carefully distinguish between the use of a Latin square to provide the design for the randomization procedure and the use of a Latin square for other purposes.

In biological experimentation, the Latin square provides a method for controlling individual differences among experimental units. For example, suppose that it is desired to control differences among litters as well as differences among sizes within litter for animals assigned to given treatment conditions. If there are four treatments, and if the size of the litter is four, then a 4×4 Latin square may be used to obtain the

required balance. Use of this plan provides the experimenter with a dual balance.

Litter	Size within litter			
	1	2	3	4
1	t_2	t_1	t_4	t_3
2	t_3	t_4	t_1	t_2
3	t_1	t_2	t_3	t_4
4	t_4	t_3	t_2	t_1

That is, the experimental units under any treatment are balanced with respect to both litter and size within litter. This plan may be replicated as many times as is required. An alternative plan, not having this dual balance, would assign the experimental units at random to the treatment conditions. This plan is to be preferred to the balanced plan when the variables under control by restricted randomization do not in fact reduce the experimental error sufficiently to offset the loss in the degrees of freedom for estimating the experimental error.

In the behavioral sciences, the more usual application of the Latin square is in selecting a balanced fractional replication from a complete factorial experiment. To illustrate this procedure, consider an experiment conducted in accordance with the following plan (in this plan c_1 , c_2 , and c_3 represent three categories of patients):

	Drug 1	Drug 2	Drug 3
Hospital 1	c_2	c_1	c_3
Hospital 2	c_1	c_3	c_2
Hospital 3	c_3	c_2	c_1

Suppose that 15 patients are sampled from each of the hospitals, 5 patients from each of the three categories of patients. According to this design:

Patients from hospital 1 in category 2 are given drug 1.

Patients from hospital 1 in category 1 are given drug 2.

Patients from hospital 1 in category 3 are given drug 3.

Upon completion of the experiment, drug 1 is administered to some patients from each of the categories as well as some patients from each of the hospitals. This same kind of balance also holds for the other drugs.

In order to have five observations in each cell, the complete factorial would require 45 patients from each hospital, 15 patients in each category from each of the hospitals, that is, 135 patients in all. The Latin-square design requires a total of 45 patients. In this context, a Latin square represents one-third of the complete factorial experiment. Under some conditions, however, the Latin-square design may provide the experimenter

with all the important information. If interactions with the hospital factor may be considered negligible relative to the hospital main effects and the drug by category interaction, then this design provides the experimenter with complete information on the main effects of drugs and categories as well as partial information on the main effects of hospitals and drug by category interaction. Prior experimentation in an area and familiarity with the subject matter are the best guides as to whether assumptions about negligible interactions are warranted. Pilot studies are also a valuable source of information about underlying assumptions.

Another important use of the Latin square in the area of the behavioral sciences is to counterbalance order effects in plans calling for repeated measures. For example, consider the following experimental plan:

	Order 1	Order 2	Order 3	Order 4
Group 1	a_3	a_1	a_2	a_4
Group 2	a_2	a_3	a_4	a_1
Group 3	a_4	a_2	a_1	a_3
Group 4	a_1	a_4	a_3	a_2

Suppose that each of the groups represents a random subsample from a larger random sample. Each of the subjects within group 1 is given treatments a_1 through a_4 in the order indicated by the columns of the Latin square. Upon completion of the experiment, each of the levels of factor A will be administered once in each of the orders. Hence there is a kind of balance with respect to the order effect. If an order effect exists, it is "controlled" in the weak sense that each treatment appears equally often in each of the orders. Implicit in this type of "control" are the assumptions of a strictly additive model with respect to the order factor. That is, the order factor is assumed to be additive with respect to the treatment and group factors; also, order is assumed not to interact with the latter factors. (Further, the homogeneity-of-covariance assumptions underlying a repeated-measure design are still required for the validity of the final tests.)

In another sense, however, order effects are not under control. Consider the administration of a_2 under order 3. For the plan given above, a_2 is preceded by a_3 and a_1 in that order. Should the sequence in which a_1 and a_3 precede a_2 have any appreciable effect, this plan does not provide an adequate control on such sequence (carry-over) effects. (The term sequence effect will be used to indicate differential effects associated with the order in which treatments precede other treatments.) Differences among the groups in this plan essentially represent differences among the four sequences called for by this plan.

A modification of the plan that has just been considered leads to another important use of the Latin square. Suppose that each of the groups is

assigned to a different level of factor B . If the sequence effects are negligible relative to the effects of factor B , differences among the groups measure the effects due to factor B . If interactions with order effects are negligible, then partial information with respect to the AB interaction may be obtained. Assuming n subjects in each of the groups, the analysis of this plan takes the following form:

Source of variation	df
Between subjects	$4n - 1$
Groups (B)	3
Subjects within groups	$4(n - 1)$
Within subjects	$12n$
Order	3
A	3
$(AB)'$ (partial information)	6
Residual	$12n - 12$

In many applications, the Latin square forms a building block within a more comprehensive design. Use of a Latin square provides economy of experimental effort only in cases where certain of the interactions (to be specified more completely in the following section) are zero. Complete factorial experiments result in unnecessary experimental effort in areas where the assumptions underlying the analysis of Latin squares are appropriate.

10.5 Analysis of Latin-square Designs—No Repeated Measures

Plan 1. Consider the following 3×3 Latin square as a fractional replication of a $3 \times 3 \times 3$ factorial experiment:

	b_1	b_2	b_3
a_1	c_2	c_1	c_3
a_2	c_3	c_2	c_1
a_3	c_1	c_3	c_2

Assume there are n observations in each cell. Assume further that two-factor and three-factor interactions are negligible relative to main effects. The model for an observation made in cell ijk may be expressed as

$$X_{ijkm} = \mu + \alpha_{i(s)} + \beta_{j(s)} + \gamma_{k(s)} + \text{res}_{(s)} + \varepsilon_{m(ijk)}.$$

The subscript (s) is used to indicate that the effect in question is estimated from data obtained from a Latin square. The term $\text{res}_{(s)}$ includes all sources of variation due to treatment effects which are not predictable from

the sum of the main effects. Under this model, the analysis and the expected values of the mean square are outlined below:

Source of variation	df	df for general case	E(MS)
<i>A</i>	2	$p - 1$	$\sigma_e^2 + np\sigma_\alpha^2$
<i>B</i>	2	$p - 1$	$\sigma_e^2 + np\sigma_\beta^2$
<i>C</i>	2	$p - 1$	$\sigma_e^2 + np\sigma_\gamma^2$
Residual	2	$(p - 1)(p - 2)$	$\sigma_e^2 + \sigma_{\text{res}}^2$
Within cell	$9(n - 1)$	$p^2(n - 1)$	σ_e^2

If, indeed, interactions are negligible, the variance due to the residual sources should not differ appreciably from the variance due to experimental error. According to the expected values of the mean squares, when $\sigma_{\text{res}}^2 = 0$, the expected values for the residual and the within-cell variances are both estimates of variance due to experimental error. A partial test on the appropriateness of the model is therefore given by the F ratio

$$F = \frac{MS_{\text{res}}}{MS_{\text{w. cell}}}.$$

The magnitude of this F ratio indicates the extent to which the observed data conform to the model (which postulates no interactions). However, a test of this kind is somewhat unsatisfactory. Decisions as to the appropriateness of the model should, in general, be based upon evidence independent of that obtained in the actual experiment. Pilot studies, guided by subject-matter knowledge, should, wherever possible, serve as the basis for formulating the model under which the analysis is to be made.

Granting the appropriateness of the model, tests on main effects can readily be carried out. If interaction is present, some of the tests on main effects will no longer be possible. If, for example, all interactions with factor A may be considered negligible but the BC interaction not negligible, then the main effect of factor A will no longer be clear of interaction effects. However, main effects of B and C will be clear of interaction effects; i.e., tests on these latter main effects can be carried out.

Numerical Example. Computational procedures for the plan to be described are similar to those used in a two-factor factorial experiment. The data in Table 10.5-1 will be used for illustrative purposes. Suppose that an experimenter is interested in evaluating the relative effectiveness of three drugs (factor B) on three categories (factor C) of patients. Patients for the experiment are obtained from three different hospitals (factor A).

The experimenter obtains 12 patients from each of the hospitals. Of these 12 patients, 4 belong to category c_1 , 4 to category c_2 , and 4 to category c_3 . The total number of patients in the experiment is 36; the total number of patients in any one category is 12.

The design for the experiment is the Latin square given at the left of part i. The first row of this Latin square indicates the manner in which the 12 patients from hospital a_2 are assigned to the drugs. The four patients in category c_3 are given drug b_1 ; the four patients in category c_2 are given drug b_3 ; the four patients in category c_1 are given drug b_2 . The criterion

Table 10.5-1 Numerical Example

Design:

	b_1	b_3	b_2
a_2	c_3	c_2	c_1
a_1	c_2	c_1	c_3
a_3	c_1	c_3	c_2

Observed data:

	b_1	b_3	b_2
a_2	6, 8, 12, 7	0, 0, 1, 4	0, 2, 2, 5
a_1	2, 5, 3, 1	2, 2, 4, 6	9, 10, 12, 12
a_3	0, 1, 1, 4	2, 1, 1, 5	0, 1, 1, 4

(i)

Cell totals

	b_1	b_3	b_2	Total	
a_2	33	5	9	$47 = A_2$	$n = 4$ $p = 3$
a_1	11	14	43	$68 = A_1$	
a_3	6	9	6	$21 = A_3$	
Total	50	28	58	$136 = G$	
	B_1	B_3	B_2		
	29	22	85		
	C_1	C_2	C_3		

(ii)

<p>(1) = $G^2/np^2 = 513.78$</p> <p>(2) = $\sum X^2 = 978.00$</p> <p>(3) = $\sum A^2/np = 606.17$</p>	<p>(4) = $\sum B^2/np = 554.00$</p> <p>(5) = $\sum C^2/np = 712.50$</p> <p>(6) = $\sum (ABC)^2/n = 878.50$</p>
--	---

(iii)

<p>$SS_a = (3) - (1) = 92.39$</p> <p>$SS_b = (4) - (1) = 40.22$</p> <p>$SS_c = (5) - (1) = 198.72$</p> <p>$SS_{res} = (6) - (3) - (4) - (5) + 2(1) = 33.39$</p>	<p>$SS_{\text{between cells}} = (6) - (1) = 364.72$</p> <p>$SS_{\text{w. cells}} = (2) - (6) = 99.50$</p>
---	---

scores for these three sets of four patients are given in the first row of the table of observed data.

In part ii are the cell totals for the four observations within each cell of the design. The entry in row a_2 , column b_1 , which is 33, is the sum of the criterion scores for the four patients in category c_3 from hospital a_2 —these four patients were given drug b_1 . This total will be designated by the symbol ABC_{213} . Similarly, the entry in row a_1 , column b_2 , which is 43, will be designated by the symbol ABC_{123} —this total is the sum of the four observations under treatment conditions abc_{123} . Row totals and column totals in part ii define, respectively, the A_i 's and the B_j 's. The C_k 's are also obtained

from the cell totals. From the design, one locates cells at specified levels of factor C ; thus,

$$C_1 = 9 + 14 + 6 = 29,$$

$$C_2 = 5 + 11 + 6 = 22,$$

$$C_3 = 33 + 43 + 9 = 85.$$

Computational symbols are defined in part iii. In terms of the symbols, the variation due to the main effect of factor A is estimated by

$$SS_a = np\Sigma(\bar{A}_i - \bar{G})^2 = (3) - (1) = 92.39.$$

The variation due to the main effect of factor B is estimated by

$$SS_b = np\Sigma(\bar{B}_j - \bar{G})^2 = (4) - (1) = 40.22.$$

Similarly, the variation due to the main effect of factor C is estimated by

$$SS_c = np\Sigma(\bar{C}_k - \bar{G})^2 = (5) - (1) = 198.72.$$

The residual sum of squares is computed in a somewhat indirect manner. The variation between the nine cells in the Latin square is

$$SS_{\text{between cells}} = n\Sigma(\overline{ABC}_{ijk} - \bar{G})^2 = (6) - (1) = 364.72.$$

This source of variation represents a composite of all factorial effects, main effects as well as interactions. There are eight $p^2 - 1$ degrees of freedom in this source of variation. The residual variation is given by

$$\begin{aligned} SS_{\text{res}} &= SS_{\text{between cells}} - SS_a - SS_b - SS_c \\ &= (2) - (3) - (4) - (5) + 2(1) = 33.39. \end{aligned}$$

From this point of view, the residual is that part of the between-cell variation which cannot be accounted for by additivity of the three main effects. If no interactions exist, this source of variation provides an estimate of experimental error. If interactions do exist, the residual variation is, in part, an estimate of these interactions.

The within-cell variation is estimated by

$$SS_{\text{w. cell}} = \sum_m (X_{ijkm} - \overline{ABC}_{ijk})^2 = (2) - (6) = 99.50.$$

A summary of the analysis of variance is given in Table 10.5-2.

A partial test of the hypothesis that all interactions are negligible is given by

$$F = \frac{MS_{\text{res}}}{MS_{\text{w. cell}}} = \frac{16.70}{3.69} = 4.53.$$

For a .05-level test, the critical value is $F_{.05}(2,27) = 3.35$. Hence the experimental data tend to contradict the hypothesis that the interactions are negligible. Under these conditions, the adequacy of the Latin-square design is questionable—estimates of the main effects may be confounded by interaction terms.

If all interactions with the hospital factor are negligible, then the variation due to the residual represents partial information on the drug by category interaction. Under this latter assumption, the main effects of the drug factor and category factor will not be confounded with interaction effects. If the assumption is made that all interactions with the hospital factor are negligible, the experimental data indicate that the drug by category interaction is statistically significant. Tests on main effects for drugs and categories also indicate statistically significant variation.

Table 10.5-2 Analysis of Variance for Numerical Example

Source of variation	SS	df	MS	F
Hospitals (A)	92.39	2	46.20	12.52*
Drugs (B)	40.22	2	20.11	5.45*
Categories (C)	198.72	2	99.36	26.93*
Residual	33.39	2	16.70	4.53*
Within cell	99.50	27	3.69	
Total	464.22	35		

$$* F_{.95}(2,27) = 3.35$$

The test on the main effects due to hospitals cannot adequately be made since this source of variation is partially confounded with the drug by category interaction. If the hospital factor is a random factor, there will ordinarily be little intrinsic interest in this test.

The experimental data do indicate statistically significant differences between the relative effectiveness of the drugs; however, in view of the presence of drug by category interaction, the interpretation of these differences requires an analysis of the simple effects. Information for this latter type of analysis is not readily obtained from a Latin-square design. Had the existence of an interaction effect been anticipated, a complete factorial experiment with respect to the interacting factors would have been the more adequate design.

To show the relationship between the analysis of variance for a Latin square and the analysis for a two-factor factorial experiment, suppose factor A in Table 10.5-1 is disregarded. Then the analysis of variance is as follows:

Source	SS	df
B	(4) - (1) = 40.22	2
C	(5) - (1) = 198.72	2
BC	$\frac{\Sigma(BC)^2}{n} - (4) - (5) + (1) = 125.78$	4
Within cell	(2) - (6) = 99.50	27

The two degrees of freedom for the main effect of factor A are part of the BC interaction of the two-factor factorial experiment. Thus, in a Latin square the interaction term is partitioned as follows:

BC	SS	df
A	92.39	2
Residual	$SS_{bc} - SS_a = 33.39$	2
	125.78	4

Plan 2. Use of a Latin square as part of a larger design is illustrated in the following plan:

		b_1	b_2	b_3
d_1	a_1	c_3	c_2	c_1
	a_2	c_1	c_3	c_2
	a_3	c_2	c_1	c_3

		b_1	b_2	b_3
d_2	a_1	c_1	c_3	c_2
	a_2	c_2	c_1	c_3
	a_3	c_3	c_2	c_1

Assume that there are n independent observations within each cell. All observations in a single square are made at the same level of factor D . The separate squares are at different levels of factor D . In some plans it is desirable to choose a balanced set of squares, in others use of the same square throughout is to be preferred, while in still other plans independent randomization of the squares is to be preferred. Guides in choosing the squares will be discussed after the analysis has been outlined.

A model for which this plan provides adequate data is the following:

$$E(X_{ijkmo}) = \mu + \alpha_{i(s)} + \beta_{j(s)} + \gamma_{k(s)} + \delta_m + \alpha\delta_{i(s)m} + \beta\delta_{j(s)m} + \gamma\delta_{k(s)m} + \text{res}_{(s)}.$$

The subscript (s) indicates that an effect forms one of the dimensions of the square. It will be noted that no interactions are assumed to exist between factors that form part of the same Latin square (i.e., factors A , B , and C). However, interactions between the factor assigned to the whole square (factor D) and the factors that form the parts are included in the model.

Assuming that all factors are fixed and that there are n observations in each cell, the appropriate analysis and expected values for the mean squares (as obtained from the above model) are outlined in Table 10.5-3. A partial check on the appropriateness of the model is provided by the F ratio

$$F = \frac{MS_{\text{res}}}{MS_{\text{w. cell}}}.$$

When the model is appropriate, this ratio should not differ appreciably from 1.00.

Use of this experimental plan requires a highly restrictive set of assumptions with respect to some of the interactions. Should these assumptions be violated, main effects will be partially confounded with interaction effects.

If, for example, it may be assumed that interactions with factor A are negligible, but that the BC interaction cannot be considered negligible, then the main effect of factor A will be confounded with the BC interaction. However, main effects of factors B and C will not be confounded. When interactions with factor A are negligible, partial information on the BC interaction is available from the within-square residuals. When such information is to be used to make inferences about the BC interaction, it is generally advisable (if the design permits) to work with a balanced set of Latin squares rather than to work with independently randomized squares.

Table 10.5-3 Analysis of Plan 2

Source of variation	df	df for general case	E(MS)
A	2	$p - 1$	$\sigma_\varepsilon^2 + npq\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_\varepsilon^2 + npq\sigma_\beta^2$
C	2	$p - 1$	$\sigma_\varepsilon^2 + npq\sigma_\gamma^2$
D	1	$q - 1$	$\sigma_\varepsilon^2 + np^2\sigma_\delta^2$
AD	2	$(p - 1)(q - 1)$	$\sigma_\varepsilon^2 + np\sigma_{\alpha\delta}^2$
BD	2	$(p - 1)(q - 1)$	$\sigma_\varepsilon^2 + np\sigma_{\beta\delta}^2$
CD	2	$(p - 1)(q - 1)$	$\sigma_\varepsilon^2 + np\sigma_{\gamma\delta}^2$
Residual	4	$q(p - 1)(p - 2)$	$\sigma_\varepsilon^2 + n\sigma_{es}$
Within cell	$18(n - 1)$	$p^2q(n - 1)$	σ_ε^2

Use of a balanced set of squares will provide partial information on all the components of the interaction term. Use of the same square throughout is to be avoided in this context, since information on only a limited number of the interaction components will be available.

Computational Procedures for Plan 2. The computational procedures assume that there are q levels of factor D and n observations in each cell of the $p \times p$ Latin squares included in the plan. The observed data will consist of q squares, one for each level of factor D ; each square will contain np^2 observations. Hence the total number of observations in the experiment is np^2q . From the observed data, a summary table of the following form may be prepared (for illustrative purposes, assume that $p = 3$ and $q = 2$):

	d_1	d_2	
a_1	AD_{11}	AD_{12}	A_1
a_2	AD_{21}	AD_{22}	A_2
a_3	AD_{31}	AD_{32}	A_3
	D_1	D_2	G

Each of the AD_{im} 's in this summary table is the sum of the np observations in the experiment which were made under treatment combination ad_{im} .

From this summary table, one may compute the following sums of squares:

$$SS_a = \frac{\sum A_i^2}{npq} - \frac{G^2}{np^2q},$$

$$SS_d = \frac{\sum D_m^2}{np^2} - \frac{G^2}{np^2q},$$

$$SS_{ad} = \frac{\sum (AD_{im})^2}{np} - \frac{\sum A_i^2}{npq} - \frac{\sum D_m^2}{np^2} + \frac{G^2}{np^2q}.$$

In a manner analogous to that by which the AD summary table is constructed, one may construct a BD summary table. From this latter table one computes SS_b and SS_{bd} . From a CD summary table one can compute SS_c and SS_{cd} .

To obtain the variation due to the residual sources, one proceeds as follows: The Latin square assigned to level d_m will be called square m . The residual variation within square m is given by

$$SS_{\text{res}(m)} = \frac{\sum_{(m)} (\text{cell totals})^2}{n} - \frac{\sum_{(m)} A_i^2}{np} - \frac{\sum_{(m)} B_j^2}{np} - \frac{\sum_{(m)} C^2}{np} + \frac{2G_m^2}{np^2}.$$

The notation $\sum_{(m)}$ indicates that the summation is restricted to square m . The degrees of freedom for residual variation within square m are

$$(p-1)(p-2).$$

The residual variation for the whole experiment is

$$SS_{\text{res}} = \sum SS_{\text{res}(m)}.$$

The degrees of freedom for SS_{res} are

$$df_{\text{res}} = \sum (p-1)(p-2) = q(p-1)(p-2).$$

As a partial check on the assumptions underlying the analysis, the within-square residuals for each of the squares should be homogeneous. One may use an F_{max} test for this purpose.

The within-cell variation is given by

$$SS_{\text{w. cell}} = \sum X^2 - \frac{\sum (\text{cell total})^2}{n},$$

where the summation is over the entire experiment. To check the assumption of homogeneity of error variation, the summation in the last expression may be restricted to the individual cells. The qp^2 sums of squares obtained in this manner may then be checked for homogeneity by means of an F_{max} test.

As an additional check on the assumption of no interactions between the factors which form the dimensions of the squares, one has the F ratio

$$F = \frac{MS_{\text{res}}}{MS_{\text{w. cell}}}$$

When $\sigma_{\text{res}}^2 = 0$, this F ratio should be approximately unity. The residual variation is sometimes pooled with the within-cell variation to provide a pooled estimate of the experimental error, i.e.,

$$SS_{\text{pooled error}} = SS_{\text{res}} + SS_{\text{w. cell.}}$$

The degrees of freedom for $SS_{\text{pooled error}}$ is the sum of the degrees of freedom for the parts.

Plan 3. This plan resembles Plan 2 in form, but the assumptions underlying it are quite different. Here the treatment combinations in a $p \times p \times p$ factorial experiment are divided into a balanced set of $p \times p$ Latin squares. The levels of factors A , B , and C are then assigned at random to the symbols defining the Latin square. Then the levels of factor D are assigned at random to the whole squares. Hence the number of levels for each factor must be p . An illustration in terms of a balanced set of 3×3 Latin squares is given below:

		b_1	b_2	b_3			b_1	b_2	b_3			b_1	b_2	b_3
d_1	a_1	c_1	c_2	c_3	d_2	a_1	c_2	c_3	c_1	d_3	a_1	c_3	c_1	c_2
	a_2	c_2	c_3	c_1		a_2	c_3	c_1	c_2		a_2	c_1	c_2	c_3
	a_3	c_3	c_1	c_2		a_3	c_1	c_2	c_3		a_3	c_2	c_3	c_1

In this plan all the 27 possible combinations of the factors A , B , and C are present. However, only one-third of the 81 possible combinations of the factors A , B , C , and D that would be present in the complete four-factor factorial experiment are present in the above plan. If interactions with factor D are negligible, then this plan will provide complete information with respect to the main effects of factors A , B , and C , as well as all two-factor interactions between these factors. The main effect of factor D will be partially confounded with the ABC interaction; however, partial information with respect to the ABC interaction may be obtained.

This design is particularly useful when the experimenter is interested primarily in a three-factor factorial experiment in which control with respect to the main effects of a fourth factor is deemed desirable. The model appropriate for this plan is the following:

$$E(X_{ijkmo}) = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \delta_m + \alpha\beta\gamma'_{ijk}$$

The prime symbol on the three-factor interaction indicates only partial information. (This plan can perhaps be more accurately classified as a balanced incomplete-block design.) Note that no interactions with factor D are included in the model; should such interactions exist, estimates of the

sources of variation due to factors A , B , C , and their two-factor interactions will be confounded with interactions with factor D .

Assuming (1) that the model is appropriate for the experimental data, (2) that there are n observations in each of the p^3 cells in the experiment, and (3) that A , B , and C are fixed factors, an outline of the analysis and the expected values of the mean squares are given in Table 10.5-4.

Table 10.5-4 Analysis of Plan 3

Source of variation	df	df for general case	E(MS)
A	2	$p - 1$	$\sigma_\varepsilon^2 + np^2\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_\varepsilon^2 + np^2\sigma_\beta^2$
C	2	$p - 1$	$\sigma_\varepsilon^2 + np^2\sigma_\gamma^2$
AB	4	$(p - 1)^2$	$\sigma_\varepsilon^2 + np\sigma_{\alpha\beta}^2$
AC	4	$(p - 1)^2$	$\sigma_\varepsilon^2 + np\sigma_{\alpha\gamma}^2$
BC	4	$(p - 1)^2$	$\sigma_\varepsilon^2 + np\sigma_{\beta\gamma}^2$
D	2	$(p - 1)$	$\sigma_\varepsilon^2 + np^2\sigma_\delta^2$
$(ABC)'$	6	$(p - 1)^3 - (p - 1)$	$\sigma_\varepsilon^2 + n\sigma_{\alpha\beta\gamma}^2$
Within cell	$27(n - 1)$	$p^3(n - 1)$	σ_ε^2

Computational Procedures for Plan 3. Because this plan includes a balanced set of Latin squares, one may construct an ABC summary table of the following form:

	b_1			b_2			b_3		
	c_1	c_2	c_3	c_1	c_2	c_3	c_1	c_2	c_3
a_1	ABC_{111}			ABC_{121}			ABC_{131}		
a_2									
a_3									

Each total of the form ABC_{ijk} is based upon n observations. By means of the usual computational formulas for a three-factor factorial experiment having n observations per cell, one may obtain the sums of squares for all the factorial effects, including SS_{abc} . For example,

$$SS_a = \frac{\sum A_i^2}{np^2} - \frac{G^2}{np^3}.$$

The sum of squares due to the three-factor interaction, SS_{abc} , includes the variation due to the main effect of factor D . The adjusted sum of squares for this three-factor interaction is given by

$$SS'_{abc} = SS_{abc} - \left(\frac{\sum D_m^2}{np^2} - \frac{G^2}{np^3} \right).$$

SS'_{abc} includes that part of the ABC interaction which is not confounded with the main effect due to factor D . The degrees of freedom for SS'_{abc} are

$$\begin{aligned} df_{SS'_{abc}} &= df_{SS_{abc}} - df_{SS_d} \\ &= (p-1)^3 - (p-1). \end{aligned}$$

The within-cell variation is given by

$$SS_{w, cell} = \sum X^2 - \frac{\sum(\text{cell total})^2}{n},$$

where the notation (cell total) represents the sum of the n observations in the experiment made under a unique combination of factors A , B , C , and D . The summation is over all cells in the experiment.

Plan 4. This plan uses the treatment combinations making up a factorial set as one or more dimensions of a Latin square. From many points of view, Plan 4 may be regarded as a special case of Plan 1. A square of dimension $p^2 \times p^2$ is required for this plan. It will be illustrated by the following 4×4 square:

	cd_{11}	cd_{12}	cd_{21}	cd_{22}
ab_{11}	t_2	t_1	t_3	t_4
ab_{12}	t_3	t_2	t_4	t_1
ab_{21}	t_1	t_4	t_2	t_3
ab_{22}	t_4	t_3	t_1	t_2

The treatment combinations along the rows are those forming a 2×2 factorial experiment; the treatment combinations across the columns also form a 2×2 factorial experiment. The letters in the square may or may not form a factorial set. For illustrative purposes suppose that they do not. Assume that there are n observations in each cell. If the treatment sets which define the dimensions of the Latin square do not interact, then an appropriate analysis is given in Table 10.5-5. The expected values of the mean squares assume that A , B , C , and D are fixed factors. (Note that the factors within a dimension of the Latin square may interact.)

Computational Procedures for Plan 4. To obtain SS_a , SS_b , and SS_{ab} , the p^2 row totals are arranged in the form of an AB summary table.

	b_1	b_2	
a_1	AB_{11}	AB_{12}	A_1
a_2	AB_{21}	AB_{22}	A_2
	B_1	B_2	G

Each AB_{ij} is the sum of np^2 observations. Hence each A_i total is based upon np^3 observations. The variation due to the main effects of factor A is

$$SS_a = \frac{\sum A_i^2}{np^3} - \frac{G^2}{np^4}.$$

The variation due to the AB interaction is

$$SS_{ab} = \frac{\Sigma(AB_{ij})^2}{np^2} - \frac{\Sigma A_i^2}{np^3} - \frac{\Sigma B_j^2}{np^3} + \frac{G^2}{np^4}.$$

The sums of squares SS_c , SS_d , and SS_{cd} are computed in an analogous manner from the column totals.

Table 10.5-5 Analysis of Plan 4

Source of variation	df	df for general case	E(MS)
Row effects	3	$p^2 - 1$	
<i>A</i>	1	$p - 1$	$\sigma_e^2 + np^2\sigma_\alpha^2$
<i>B</i>	1	$p - 1$	$\sigma_e^2 + np^2\sigma_\beta^2$
<i>AB</i>	1	$(p - 1)^2$	$\sigma_e^2 + np\sigma_{\alpha\beta}^2$
Column effects	3	$p^2 - 1$	
<i>C</i>	1	$p - 1$	$\sigma_e^2 + np^2\sigma_\gamma^2$
<i>D</i>	1	$p - 1$	$\sigma_e^2 + np^2\sigma_\delta^2$
<i>CD</i>	1	$(p - 1)^2$	$\sigma_e^2 + np\sigma_{\gamma\delta}^2$
Letters in cells	3	$p^2 - 1$	
<i>T</i>	3	$p^2 - 1$	$\sigma_e^2 + np^2\sigma_\tau^2$
Residual	6	$(p^2 - 1)(p^2 - 2)$	$\sigma_e^2 + n\sigma_{\text{res}}^2$
Within cell	$16(n - 1)$	$p^4(n - 1)$	σ_e^2

If the sum of the np^2 observations at level t_o is designated by the symbol T_o , then the variation due to the treatments assigned to the letters of the square is

$$SS_t = \frac{\Sigma T_o^2}{np^2} - \frac{G^2}{np^4}.$$

The variation due to residual sources is

$$SS_{\text{res}} = \frac{\Sigma(\text{cell total})^2}{n} - \frac{\Sigma(AB_{ij})^2}{np^2} - \frac{\Sigma(CD_{km})^2}{np^2} - \frac{\Sigma T_o^2}{np^2} + \frac{2G^2}{np^4}.$$

The within-cell variation is

$$SS_{\text{w. cell}} = \Sigma X^2 - \frac{\Sigma(\text{cell total})^2}{n},$$

where the summation is over all cells in the experiment.

Summary of Plans in Sec. 10.5

Plan 1				Plan 2			
	b_1	b_2	b_3		b_1	b_2	b_3
a_1	c_2	c_1	c_3				
a_2	c_3	c_2	c_1				
a_3	c_1	c_3	c_2				

Plan 3				Plan 4			
	b_1	b_2	b_3		b_1	b_2	b_3
d_1	a_1	c_1	c_2	cd_{11}	t_2	t_1	t_3
a_2	c_2	c_3	c_1	cd_{12}	t_3	t_2	t_4
a_3	c_3	c_1	c_2	cd_{21}	t_1	t_4	t_2
				cd_{22}	t_4	t_3	t_1

10.6 Analysis of Greco-Latin Squares

In its classical context, a Latin-square arrangement permits a two-way control in variation of the experimental units, i.e., control of row and column effects. In a similar context, a Greco-Latin square permits a three-way control in the variation of the experimental units, i.e., row effects, column effects, and "layer" effects. Thus, a Greco-Latin square defines a restricted randomization procedure whereby p treatments are assigned to p^2 experimental units so as to obtain balance along three dimensions.

From the point of view of construction, a Greco-Latin square is a composite of two orthogonal Latin squares. (Independent randomization of rows and columns is required before the resulting composite square is used in practice.) In order to maintain the identity of the squares which are combined to form the composite, the cells of one square are often designated by Latin letters, and the cells of the second square are often designated by Greek letters. This procedure is demonstrated for two 3×3 squares.

I			II			Composite		
a	b	c	α	β	γ	$a\alpha$	$b\beta$	$c\gamma$
b	c	a	γ	α	β	$b\gamma$	$c\alpha$	$a\beta$
c	a	b	β	γ	α	$c\beta$	$a\gamma$	$b\alpha$

An equivalent representation in terms of numbers rather than letters is given below. In the resulting composite square the first digit in a pair

represents the level of one effect, and the second digit the level of a second effect.

I			II			Composite		
1	2	3	1	2	3	11	22	33
2	3	1	3	1	2	23	31	12
3	1	2	2	3	1	32	13	21

In a Greco-Latin square, there are in reality four variables—namely, row, column, Latin-letter, and Greek-letter variables. From this point of view, a Greco-Latin square may be regarded as a kind of four-factor experiment. There are p^2 cells in this square; hence there are p^2 treatment combinations. In a four-factor factorial experiment in which each factor has p levels, there are p^4 treatment combinations. A Greco-Latin square may also be regarded as a p^2/p^4 or $1/p^2$ fractional replication of a $p \times p \times p \times p$ factorial experiment.

Table 10.6-1 Analysis of Variance for Greco-Latin Square

Source of variation	df	df for general case	E(MS)
<i>A</i> (rows)	2	$p - 1$	$\sigma_e^2 + np\sigma_\alpha^2$
<i>B</i> (columns)	2	$p - 1$	$\sigma_e^2 + np\sigma_\beta^2$
<i>C</i> (Latin letters)	2	$p - 1$	$\sigma_e^2 + np\sigma_\gamma^2$
<i>D</i> (Greek letters)	2	$p - 1$	$\sigma_e^2 + np\sigma_\delta^2$
Residual	—	$(p - 1)(p - 3)$	$\sigma_e^2 + n\sigma_{\text{res}}^2$
Within cell	$9(n - 1)$	$p^2(n - 1)$	σ_e^2
Total	$9n - 1$	$np^2 - 1$	

As a fractional replication of a factorial experiment, main effects of each of the factors will be confounded with two-factor and higher-order interaction effects. For example, the main effects of factor *A* will be confounded with the *BC*, *CD*, and *BCD* interactions. In general, the utility of a single Greco-Latin square is limited to experimental situations in which the four dimensions of the square have negligible interactions. However, Greco-Latin squares may be used to good advantage as part of more inclusive designs. The latter are illustrated in Sec. 10.7. If all interactions between factors defining the dimensions of a Greco-Latin square are negligible, then the analysis of variance takes the form given in Table 10.6-1. This analysis assumes that there are n independent observations in each of the cells. The specific degrees of freedom are for the case of a 3×3 square. For this special case, the degrees of freedom of the residual variation are zero; only for the case $p > 3$ will the residual

variation be estimable. A partial check on the assumptions made about negligible interactions is given by

$$F = \frac{MS_{\text{res}}}{MS_{\text{w. cell}}}$$

Depending upon the outcome of this test, the residual sum of squares is sometimes pooled with the within-cell variation to provide an over-all estimate of the variation due to experimental error.

Table 10.6-2 Computational Procedures for Greco-Latin Square

(i)	(1) = G^2/np^2	(5) = $(\sum C_k^2)/np$
	(2) = $\sum X^2$	(6) = $(\sum D_m^2)/np$
	(3) = $(\sum A_i^2)/np$	(7) = $[\sum (\text{cell total})^2]/n$
	(4) = $(\sum B_j^2)/np$	
(ii)	<i>A</i>	(3) - (1)
	<i>B</i>	(4) - (1)
	<i>C</i>	(5) - (1)
	<i>D</i>	(6) - (1)
	Residual	(7) - (3) - (4) - (5) - (6) + 3(1)
	Within cell	(2) - (7)
	Total	(2) - (1)

A summary of the computational procedures is given in Table 10.6-2. These procedures differ only slightly from those appropriate for the Latin square. The residual variation is obtained from the relation

$$SS_{\text{res}} = SS_{\text{between cells}} - SS_a - SS_b - SS_c - SS_d.$$

Residual variation includes interaction terms if these are not negligible. Otherwise the residual variation provides an estimate of experimental error.

10.7 Analysis of Latin Squares—Repeated Measures

In the plans that are discussed in this section, all the restrictions on the model underlying a repeated-measure design for a factorial experiment are necessary in order that the final F tests be valid. These restrictions were discussed in Chap. 7. Special attention should be given to the possible presence of nonadditive sequence effects in experiments which do not involve learning. In particular, a repeated-measure design assumes that all pairs of observations on the same subjects have a constant correlation. If this assumption is violated, resulting tests on within-subject effects tend to be biased in the direction of yielding too many significant results.

The equivalent of the conservative test proposed by Box (1954) and by Greenhouse and Geisser (1959) can be adapted for use in connection with the designs that follow. For example, if the test for a within-subject main

effect (as indicated by the model requiring the homogeneity-of-correlation condition) requires the critical value

$$F_{1-\alpha}[(p-1), p(n-1)(p-1)],$$

then the corresponding critical value under the conservative test procedure is

$$F_{1-\alpha}[1, p(n-1)].$$

In principle, the degrees of freedom for the numerator and denominator of the F distribution required in the usual test (as given by the expected values of the mean squares obtained from the restricted model) are divided by the degrees of freedom for the factor on which there are repeated measures.

Plan 5. Consider the following 3×3 Latin square in which groups of n subjects are assigned at random to the rows:

	a_1	a_2	a_3
G_1	b_3	b_1	b_2
G_2	b_1	b_2	b_3
G_3	b_2	b_3	b_1

In this plan each of the n subjects in G_1 is observed under all treatment combinations in row 1, that is, $ab_{13}, ab_{21}, ab_{32}$. For example, the levels of factor A may represent three kinds of targets, and the levels of factor B may represent three distances. There are nine possible treatment combinations in the complete 3×3 factorial experiment; each of these nine appears in the Latin square. However, the individuals within any one group are observed only under three of the nine possibilities. Suppose that the order in which a subject is observed under each treatment combination is randomized independently for each subject.

If the interactions with the group factor are negligible, the following model will be appropriate for the analysis (this assumption is reasonable if the groups represent random subsamples from a common population):

$$E(X_{ijkm}) = \mu + \delta_k + \pi_{m(k)} + \alpha_i + \beta_j + \alpha\beta'_{ij}.$$

In this model δ_k represents effects associated with the groups and $\pi_{m(k)}$ effects associated with subjects within the groups. The symbol $\alpha\beta'_{ij}$ indicates that only partial information is available on this source of variation. Assuming that factors A and B are fixed factors, the analysis and the expected values of the mean squares are given in Table 10.7-1.

In this analysis, only $(p-1)(p-2)$ degrees of freedom for the AB interaction appear as within-subject effects. The missing $p-1$ degrees of freedom define the variation among the groups. Since differences among the groups, in part, reflect differences due to the effects of various combinations of A and B (which are balanced with respect to main effects), such differences define part of the AB interaction. It is readily shown that

$$SS_{ab} = SS_{\text{groups}} + SS'_{ab},$$

where SS_{ab} is the variation due to the AB interaction as computed in a two-factor factorial experiment. From some points of view, SS_{groups} may be regarded as the between-subject component of the AB interaction. For most practical purposes, only SS'_{ab} (the within-subject component) is tested; tests on the latter component will generally be the more powerful.

Table 10.7-1 Analysis of Plan 5

Source of variation	df	df for general case	E(MS)
Between subjects	$3n - 1$	$np - 1$	
Groups	2	$p - 1$	$\sigma_\varepsilon^2 + p\sigma_\pi^2 + np\sigma_\delta^2$
Subjects within groups	$3(n - 1)$	$p(n - 1)$	$\sigma_\varepsilon^2 + p\sigma_\pi^2$
Within subjects	$6n$	$np(p - 1)$	
A	2	$p - 1$	$\sigma_\varepsilon^2 + np\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_\varepsilon^2 + np\sigma_\beta^2$
$(AB)'$	2	$(p - 1)(p - 2)$	$\sigma_\varepsilon^2 + n\sigma_{\alpha\beta}^2$
Error (within)	$6n - 6$	$p(n - 1)(p - 1)$	σ_ε^2

The appropriateness of the model should be given serious consideration before this plan is used for an experiment. A possible alternative plan, which requires the same amount of experimental effort but does not utilize the Latin-square principle, is the following:

	a_1	a_2	a_3
G_1	b_1	b_1	b_1
G_2	b_2	b_2	b_2
G_3	b_3	b_3	b_3

Again the groups are assumed to be random samples of size n from a common population. In this plan each of the n subjects in group 1 is observed under the treatment combinations ab_{11} , ab_{21} , and ab_{31} . Assume that the order in which subjects are observed under each treatment combination is randomized independently for each subject. In this plan there are repeated measures on factor A but no repeated measures on factor B . A model appropriate for estimation and tests under this plan is

$$E(X_{ijkm}) = \mu + \beta_j + \pi_{k(j)} + \alpha_i + \alpha\beta_{ij} + \alpha\pi_{ik(j)}.$$

For this model, assuming that factors A and B are fixed, the analysis is outlined in Table 10.7-2. In contrast to Plan 5, under this plan the AB interaction is not confounded with between-group differences; however, all components of the main effect of factor B are between-subject components. That is, differences between levels of factor B are simultaneously differences between groups of people. Hence use of the Latin square permits a more sensitive test on the main effects of factor B , provided that

the model under which the analysis is made is appropriate. In general the $A \times$ subjects within groups interaction under the alternative plan will be of the same order of magnitude as the error (within) term in Plan 5.

If the experimenter's primary interest is in factor A and its interaction with factor B , and if there is little interest in the main effects of factor B , use

Table 10.7-2 Analysis of Alternative to Plan 5

Source of variation	df	df for general case	E(MS)
Between subjects	$3n - 1$	$np - 1$	
B	2	$p - 1$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\beta^2$
Subjects within groups	$3(n - 1)$	$p(n - 1)$	$\sigma_e^2 + p\sigma_\pi^2$
Within subjects	$6n$	$np(p - 1)$	
A	2	$p - 1$	$\sigma_e^2 + \sigma_{\alpha\pi}^2 + np\sigma_\alpha^2$
AB	4	$(p - 1)(p - 1)$	$\sigma_e^2 + \sigma_{\alpha\pi}^2 + n\sigma_{\alpha\beta}^2$
$A \times$ subjects within groups	$6n - 6$	$p(n - 1)(p - 1)$	$\sigma_e^2 + \sigma_{\alpha\pi}^2$

of the Latin-square plan is not recommended. However, if the main effects of factor B are also of primary interest to the experimenter, use of the Latin square has much to recommend it. By suitable replication, some within-subject information may be obtained on all components of the AB interaction. Plan 5 forms a building block out of which other plans may be constructed.

Computational Procedures for Plan 5. Computational procedures for this plan are similar to those given for Plan 1. The sources of variation SS_a , SS_b , and SS_{groups} are computed in a manner analogous to those used in Plan 1 for corresponding effects. The variation due to the within-subject components of the AB interaction is given by

$$SS'_{ab} = SS_{ab} - SS_{\text{groups}},$$

where SS_{ab} is computed from an AB summary table by means of the usual computational formulas for this interaction.

The variation due to subjects within groups is given by

$$SS_{\text{subj w. groups}} = \frac{\sum P_{m(k)}^2}{p} - \frac{\sum G_k^2}{np},$$

where $P_{m(k)}$ represents the sum of the p observations on person m in group k , and where G_k represents the sum of the np observations in group k . The summation is over the whole experiment. If the summation is limited to a single group, one obtains $SS_{\text{subj w. group } k}$. There will be p such sources of variation. It is readily shown that

$$SS_{\text{subj w. groups}} = \sum SS_{\text{subj w. group } k}.$$

That is, the sum of squares on the left represents a pooling of the sources of variation on the right; the latter may be checked for homogeneity by means of an F_{\max} test.

It is convenient to compute $SS_{\text{error}(\text{within})}$ from $SS_{\text{w. cell}}$. The latter is given by

$$SS_{\text{w. cell}} = \Sigma X^2 - \frac{\Sigma(AB_{ij})^2}{n},$$

where AB_{ij} represents a cell total. The summation is over the whole experiment. From this source of variation,

$$SS_{\text{error}(\text{within})} = SS_{\text{w. cell}} - SS_{\text{subj w. groups}}.$$

A somewhat modified computational procedure provides the parts of the error term that may be checked for homogeneity. One first computes

$$SS_{\text{w. cell for } G_k} = \Sigma X^2 - \frac{\Sigma(AB_{ij})^2}{n},$$

where the summation is limited to those cells in which group k participates. The degrees of freedom for this source of variation are $p(n-1)$. Then

$$SS_{\text{error}(\text{within}) \text{ for } G_k} = SS_{\text{w. cell for } G_k} - SS_{\text{subj w. } G_k}.$$

Degrees of freedom for this source of variation are $p(n-1) - (n-1)$, which is equal to $(n-1)(p-1)$. It is readily shown that

$$SS_{\text{error}(\text{within})} = \Sigma SS_{\text{error}(\text{within}) \text{ for } G_k};$$

that is, the variation on the left is a pooling of the sources of variation on the right. The degrees of freedom for the pooled error (within) and the degrees of freedom of the parts are related as follows:

$$df_{\text{error}(\text{within})} = \sum_k (n-1)(p-1) = p(n-1)(p-1).$$

Thus error (within) is partitioned into p parts; each of the parts has $(n-1)(p-1)$ degrees of freedom.

Illustrative Applications. Staats et al. (1957) report an experiment which may be represented schematically as follows:

	a_1	a_2
G_1	b_1	b_2
G_2	b_2	b_1

In this experiment the levels of factor A represent two different words. The levels of factor B represent two different kinds of meaning associated with the words. Subjects were assigned at random to each of the groups. The criterion measure was a rating of the "pleasantness" of the word after a

series of associations defined by the level of factor *B*. The analysis of variance had the following form:

	df	MS	<i>F</i>
Between subjects	159		
Groups	1	0.00	
Subj w. groups	158	2.30	
Within subjects	160		
<i>A</i> Words	1	45.75	18.67
<i>B</i> Meaning	1	22.58	9.22
Error (within)	158	2.45	

In this design, differences between groups may be considered as part of the *AB* interaction.

Martindale and Lowe (1959) report an experiment which is also a special case of Plan 5. In this case, a 5×5 Latin square is central to the plan. The levels of factor *A* represent different angles for a television camera; the levels of factor *B* represent trials. There were three subjects in each of the groups. Subjects were required to track a target which was visible to them only through a television screen. The criterion was time on target. The analysis of variance had the following form:

	df	MS	<i>F</i>
Between subjects	14		
Groups	4	5,000	5.03
Subj.w. groups	10	995	
Within subjects	60		
<i>A</i> Angles	4	15,102	38.82
<i>B</i> Trials	4	167	
Pooled error	52	389	
(<i>AB</i>)'	12	597	
Error (w)	40	326	

The authors refer to (*AB*)' as the residual from the Latin square; the error (within) is called the residual within the Latin square. These latter sources of variation are pooled to form the within-subject error.

Plan 6. In Plan 5, the groups assigned to the rows of the Latin square form a quasi factor. Essentially, Plan 5 may be regarded as one complete replication of a two-factor factorial experiment arranged in incomplete blocks. From this point of view, the plan to be discussed in this section may be regarded as a fractional replication of a three-factor factorial experiment arranged in incomplete blocks.

In this plan each subject within G_1 is assigned to the treatment combinations abc_{111} , abc_{231} , and abc_{321} . Thus each subject in G_1 is observed under all levels of factors A and B but under only one level of factor C . For each subject there is balance with respect to the main effects of factors A and B , but there is no balance with respect to any of the interactions. Similarly, each subject in G_2 is assigned to the treatment combinations abc_{122} , abc_{212} ,

		a_1	a_2	a_3
G_1	c_1	b_1	b_3	b_2
G_2	c_2	b_2	b_1	b_3
G_3	c_3	b_3	b_2	b_1

and abc_{332} . Again there is balance with respect to the main effects of factors A and B , but there is no balance with respect to interactions.

Table 10.7-3 Analysis of Plan 6

Source of variation	df	df for general case	E(MS)
Between subjects	$3n - 1$	$np - 1$	
C	2	$p - 1$	$\sigma_\epsilon^2 + p\sigma_\pi^2 + np\sigma_\gamma^2$
Subjects within groups	$3(n - 1)$	$p(n - 1)$	$\sigma_\epsilon^2 + p\sigma_\pi^2$
Within subjects	$6n$	$np(p - 1)$	
A	2	$p - 1$	$\sigma_\epsilon^2 + np\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_\epsilon^2 + np\sigma_\beta^2$
Residual	2	$(p - 1)(p - 2)$	$\sigma_\epsilon^2 + n\sigma_{\text{res}}^2$
Error (within)	$6n - 6$	$p(n - 1)(p - 1)$	σ_ϵ^2

If all interactions are negligible relative to main effects (a highly restrictive assumption), the following model is appropriate for making estimates and tests in the analysis of variance:

$$E(X_{ijkm}) = \mu + \gamma_{k(s)} + \pi_{m(k)} + \alpha_{i(s)} + \beta_{j(s)} + \text{res}_{(s)}.$$

The analysis of variance and the expected values of the mean squares are summarized in Table 10.7-3.

The analysis of Plan 6 is quite similar to the analysis of Plan 5. Differences among groups in the latter plan correspond to differences due to the main effects of factor C in the present plan; what was part of the AB interaction in Plan 5 is a residual term in the present plan. It is of interest to note the relationship between the analysis of a two-factor factorial experiment which does not have repeated measures and the analysis of Plan 6. The total number of observations in Plan 6 is the same as that of a two-factor factorial experiment in which there are n observations per cell. The Latin-square arrangement in Plan 6 permits the partition of what is formally a two-factor interaction into a main effect and a residual

effect. The latter is a mixture of interaction effects if these exist. The fact that there are repeated measures in Plan 6 permits the within-cell variation to be partitioned into one part involving differences between subjects and one part which does not involve differences between subjects, provided of course that homogeneity of covariances exists.

df	Two-factor experiment	Plan 6	df
$p - 1$	A	A	$p - 1$
$p - 1$	B	B	$p - 1$
$(p - 1)(p - 1)$	AB	C	$p - 1$
		Residual	$(p - 1)(p - 2)$
$p^2(n - 1)$	Within cell	Subj w. gp	$p(n - 1)$
		Error (within)	$p(n - 1)(p - 1)$

Because of the highly restrictive assumptions with respect to the interactions, use of Plan 6 is appropriate only when experience has shown that such interactions are negligible. A partial check on this assumption is given by the ratio

$$F = \frac{MS_{\text{res}}}{MS_{\text{error(within)}}}.$$

According to the model under which the analysis of variance is made, σ_{res}^2 should be zero when the assumptions with respect to the interactions are satisfied. In cases in which one of the factors, say, A , is the experimental

Table 10.7-4 Complete Factorial Analogue of Plan 6

		a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3
G_1	c_1	b_1	b_1	b_1	b_2	b_2	b_2	b_3	b_3	b_3
G_2	c_2	b_1	b_1	b_1	b_2	b_2	b_2	b_3	b_3	b_3
G_3	c_3	b_1	b_1	b_1	b_2	b_2	b_2	b_3	b_3	b_3

variable of primary interest and factors B and C are of the nature of control factors (i.e., replications or order of presentation) this plan is potentially quite useful. With this type of experimental design interactions can frequently be assumed to be negligible relative to main effects.

In exploratory studies in which interaction effects may be of primary interest to the experimenter, there is generally no substitute for the complete factorial experiment. The complete factorial analogue of Plan 6 is represented schematically in Table 10.7-4. In order to have complete within-subject information on the main effects of factors A and B as well as complete within-subject information on all interactions, including interactions with factor C , p^2 observations on each subject are required. In Plan 6,

only p observations are made on each of the subjects. Use of Plan 6 reduces the over-all cost of the experiment in terms of experimental effort—at the possible cost, however, of inconclusive results. Should the experimenter find evidence of interaction effects which invalidate the analysis of Plan 6, he may, if his work is planned properly, enlarge the experiment into Plan 8a. It will be noted that the latter may be constructed from a series of plans having the form of Plan 6. The assumptions in Plan 8a are less restrictive with respect to interactions than are those underlying Plan 6.

Computational Procedures for Plan 6. Computational procedures for this plan are identical to those outlined for Plan 5. Here $SS_c = SS_{\text{groups}}$ and $SS'_{ab} = SS_{\text{res}}$. All other sums of squares are identical. Tests for homogeneity of error variance are also identical.

Plan 7. This plan is related to Plan 5 as well as to Plan 6. Plan 7 may be regarded as being formed from Plan 5 by superimposing an orthogonal Latin square. This plan may also be viewed as a modification of Plan 6 in which the C factor is converted into a within-subject effect. The combinations of factors B and C which appear in the cells of the following plan are defined by a Greco-Latin square:

	a_1	a_2	a_3
G_1	bc_{11}	bc_{23}	bc_{32}
G_2	bc_{22}	bc_{31}	bc_{13}
G_3	bc_{33}	bc_{12}	bc_{21}

The groups of subjects are assigned at random to the rows of the square; the levels of factor A are also assigned at random to the columns of the square. The subjects within G_1 are observed under treatment combinations abc_{111} , abc_{223} , and abc_{332} . (Assume that the order in which a subject is observed under a particular treatment combination is randomized independently for each subject.) Provided that all interactions are negligible, unbiased estimates of the differential main effects of factors A , B , and C can be obtained. Further, the expected values of the mean squares for these main effects will not involve a between-subject component. The model under which the analysis of Plan 7 can be carried out is

$$E(X_{ijkmo}) = \mu + \delta_{m(s)} + \pi_{o(m)} + \alpha_{i(s)} + \beta_{j(s)} + \gamma_{k(s)}.$$

The symbol $\delta_{m(s)}$ designates the effect of group m , and the symbol $\pi_{o(m)}$ designates the additive effect of person o within group m .

An outline of the analysis, assuming that the model is appropriate, is given in Table 10.7-5. For a 3×3 square, the variation due to the residual cannot be estimated, since there are zero degrees of freedom for this source of variation. However, for squares of higher dimension a residual term

will be estimable. The latter represents the variation due to interactions, if these exist. The variation between groups may also be regarded as part of the interaction variation.

If the levels of factor C define the order in which subjects are observed under combinations of factors A and B , then Plan 7 becomes a special case

Table 10.7-5 Analysis of Plan 7

Source of variation	df	df for general case	E(MS)
Between subjects	$3n - 1$	$np - 1$	
Groups	2	$p - 1$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\delta^2$
Subjects within groups	$3(n - 1)$	$p(n - 1)$	$\sigma_e^2 + p\sigma_\pi^2$
Within subjects	$6n$	$np(p - 1)$	
A	2	$p - 1$	$\sigma_e^2 + np\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_e^2 + np\sigma_\beta^2$
C	2	$p - 1$	$\sigma_e^2 + np\sigma_\gamma^2$
Residual	0	$(p - 1)(p - 3)$	$\sigma_e^2 + n\sigma_{\text{res}}^2$
Error (within)	$6n - 6$	$p(n - 1)(p - 1)$	σ_e^2

of Plan 5. In terms of factor C as an order factor, the schematic representation of Plan 7 given earlier in this section may be reorganized as follows:

	Order 1	Order 2	Order 3
G_1	ab_{11}	ab_{33}	ab_{22}
G_2	ab_{23}	ab_{12}	ab_{31}
G_3	ab_{32}	ab_{21}	ab_{13}

Thus the subjects within G_1 are observed under treatment combinations ab_{11} , ab_{33} , and ab_{22} in that order. With one of the factors representing an order factor, the variation between groups may be regarded as representing a sequence effect; this latter source of variation may also be regarded as part of the confounded interaction effects.

Computational Procedures for Plan 7. The computational procedures for this plan do not differ appreciably from those of the Greco-Latin square. Following the computational procedures for the latter, one obtains SS_a , SS_b , SS_c , and SS_{groups} . The between-cell variation is given by

$$SS_{\text{between cells}} = \frac{\sum(\text{cell total})^2}{n} - \frac{G^2}{np^2}.$$

From this sum of squares one obtains

$$SS_{\text{res}} = SS_{\text{between cells}} - (SS_a + SS_b + SS_c + SS_{\text{groups}}).$$

From this last relationship, the degrees of freedom for the residual source of variation are

$$\begin{aligned} df_{\text{res}} &= (p^2 - 1) - 4(p - 1) = (p - 1)(p + 1 - 4) \\ &= (p - 1)(p - 3). \end{aligned}$$

The variation due to subjects within groups is

$$SS_{\text{subj w. groups}} = \frac{\sum P_{m(k)}^2}{p} - \frac{\sum G_k^2}{np},$$

where $P_{m(k)}$ is the sum of the p observations on person m in group k , and where G_k is the sum of the np observations in group k . The error (within) variation is obtained indirectly from the within-cell variation. The latter is given by

$$SS_{\text{w. cell}} = \sum X^2 - \frac{\Sigma(\text{cell total})^2}{n},$$

where the summation is over the whole experiment. From this one obtains

$$SS_{\text{error(within)}} = SS_{\text{w. cell}} - SS_{\text{subj w. groups}}.$$

This method of computing this source of variation shows that the degrees of freedom for error (within) are

$$\begin{aligned} df_{\text{error(within)}} &= p^2(n - 1) - p(n - 1) \\ &= p(n - 1)(p - 1). \end{aligned}$$

The sum of squares for error (within) may be partitioned by the procedures given under Plan 5. That is, by limiting the summations for $SS_{\text{w. cells}}$ and $SS_{\text{subj w. groups}}$ to single groups, one may obtain p terms each of the general form $SS_{\text{error(within) for } G_k}$. The degrees of freedom for each of these terms are $(n - 1)(p - 1)$.

Illustrative Application. Leary (1958) reports a study using a plan closely related to Plan 7. This experiment may be represented schematically as follows:

Square I

Subject	Day			
	1	2	3	4
1	(11)	(22)	(33)	(44)
2	(23)	(14)	(41)	(32)
3	(34)	(43)	(12)	(21)
4	(42)	(31)	(24)	(13)

Square II

Subject	Day			
	1	2	3	4
5	(33)	(41)	(14)	(22)
6	(12)	(24)	(31)	(43)
7	(21)	(13)	(42)	(34)
8	(44)	(32)	(23)	(11)

The basic plan consists of two different Greco-Latin squares. The symbol (23), for example, represents an experimental condition having reward level

2 for list 3—a list being a set of pairs of objects. The analysis of variance for this experiment had the following form:

Between subjects		7
Squares		1
Subjects within squares		6
Within subjects		24
Days		3
Lists		3
Rewards		3
Pooled residual		15
Days × squares	3	
Lists × squares	3	
Rewards × squares	3	
Res from square I	3	
Res from square II	3	

In this design there is only one subject per cell, but there are two replications. Since the squares are considered to define a random factor, the interactions with the squares are pooled to form part of the experimental error. The residual within square I is formally equivalent to the interaction between any two of the dimensions of square I minus the sum of the main effects of the other factors in the square. Thus,

$$SS_{\text{res from I}} = SS_{\text{days} \times \text{lists}} - (SS_{\text{rewards}} + SS_{\text{subj w. I}}).$$

In each case the summations are limited to square I. This residual term is an estimate of experimental error if the assumptions underlying the model are valid. Similarly the residual from square II is an estimate of the experimental error. An F_{\max} test may be used as a partial check on the pooling of the various estimates of experimental error.

In terms of the analysis outlined in Table 10.7-5, the degrees of freedom for the pooled residual in this design are

$$(p-1)(p-3) + p(n-1)(p-1) = (3)(1) + 4(1)(3) = 15,$$

where n in this case is the number of replications.

Plan 8. This plan uses Plan 5 as a building block. Disregarding the repeated-measure aspect, this plan also resembles Plan 2. A schematic representation of Plan 8 is given below:

		I					II		
		a_1	a_2	a_3			a_1	a_2	a_3
c_1	G_1	b_1	b_2	b_3	c_2	G_4	b_2	b_3	b_1
	G_2	b_2	b_3	b_1		G_5	b_1	b_2	b_3
	G_3	b_3	b_1	b_2		G_6	b_3	b_1	b_2

In general there will be q squares, one for each level of factor C . Different squares are used for each level of factor C . In square I, all observations are at level c_1 ; in square II, all observations are at level c_2 . There is no restriction on the number of levels of factor C .

Depending upon what can be assumed about the interactions, different analyses are possible. If the interactions between factors A , B , and groups (the dimensions of the squares) are negligible, then unbiased tests can be

Table 10.7-6 Analysis of Plan 8

Source of variation	df	df for general case	E(MS)
Between subjects	$6n - 1$	$npq - 1$	
C	1	$q - 1$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\delta^2 + np^2\sigma_\gamma^2$
Groups within C	4	$q(p - 1)$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\delta^2$
Subjects within groups	$6(n - 1)$	$pq(n - 1)$	$\sigma_e^2 + p\sigma_\pi^2$
Within subjects	$12n$	$npq(p - 1)$	
A	2	$p - 1$	$\sigma_e^2 + npq\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_e^2 + npq\sigma_\beta^2$
AC	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\alpha\gamma}^2$
BC	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2$
Residual	4	$q(p - 1)(p - 2)$	$\sigma_e^2 + n\sigma_{res}^2$
Error (within)	$12n - 12$	$pq(n - 1)(p - 1)$	σ_e^2

made on all main effects and the two-factor interactions with factor C . The model under which the latter analysis is made is the following:

$$E(X_{ijkmo}) = \mu + \gamma_k + \delta_{m(k)} + \pi_{o(km)} + \alpha_{i(s)} + \beta_{j(s)} + \alpha\gamma_{ik} + \beta\gamma_{jk}.$$

The analysis outlined in Table 10.7-6 follows from this model; the expected values for the mean squares assume that factor C is fixed. It is also assumed that the order of administration of the treatment combinations is randomized independently for each subject.

The assumptions underlying Plan 8, which lead to the analysis summarized in Table 10.7-6, are different from the assumptions used in the analysis of Plan 5. A set of assumptions, consistent with those in Plan 5, leads to the analysis for Plan 8. The latter plan has the same schematic representation as Plan 8. However, in Plan 8 only the interactions involving the group factor are assumed to be negligible. Under these latter assumptions, the appropriate model is the following:

$$E(X_{ijkmo}) = \mu + \delta_k + \pi_{o(km)} + \alpha_i + \beta_j + \alpha\beta'_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk}.$$

The analysis resulting from this model, assuming factors A and B fixed, is outlined in Table 10.7-7.

Table 10.7-7 Analysis of Plan 8

Source of variation	df	df for general case	E(MS)
Between subjects	$6n - 1$	$npq - 1$	
<i>C</i>	1	$q - 1$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\delta^2 + np^2\sigma_\gamma^2$
Groups within <i>C</i>	4	$q(p - 1)$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_\delta^2$
Subjects within groups	$6(n - 1)$	$pq(n - 1)$	$\sigma_e^2 + p\sigma_\pi^2$
Within subjects	$12n$	$npq(p - 1)$	
<i>A</i>	2	$p - 1$	$\sigma_e^2 + npq\sigma_\alpha^2$
<i>B</i>	2	$p - 1$	$\sigma_e^2 + npq\sigma_\beta^2$
<i>AC</i>	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\alpha\gamma}^2$
<i>BC</i>	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2$
<i>AB'</i> from square I	2	$(p - 1)(p - 2)$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$
<i>AB'</i> from square II	2	$(p - 1)(p - 2)$	$\sigma_e^2 + n\sigma_{\alpha\beta}^2$
Error (within)	$12n - 12$	$pq(n - 1)(p - 1)$	σ_e^2

If the number of levels of factor *C* is equal to p , and if a balanced set of squares is used, then the components of the *AB* interaction which are confounded will be balanced. However, if the same Latin square is used for each of the levels of factor *C*, the same set of $p - 1$ degrees of freedom of the *AB* interaction is confounded in each square. It is of interest to compare the partition of the total degrees of freedom in a $3 \times 3 \times 3$

Table 10.7-8 Partitions in Plan 8a and $3 \times 3 \times 3$ Factorial Experiment

$3 \times 3 \times 3$ factorial		Plan 8a	
df	Source of variation	Source of variation	df
2	<i>A</i>	<i>A</i>	2
2	<i>B</i>	<i>B</i>	2
2	<i>C</i>	<i>C</i>	2
4	<i>AC</i>	<i>AC</i>	4
4	<i>BC</i>	<i>BC</i>	4
4	<i>AB</i>	(Groups within <i>C</i>	6
8	<i>ABC</i>)	($\Sigma(AB'$ for each square)	6
		(Subjects within groups	$9(n - 1)$
$27(n - 1)$	Within cell	(Error (within)	$18(n - 1)$

factorial experiment with the partition made under Plan 8a, assuming that a balanced set of squares is used in the construction of the design. These partitions are presented in Table 10.7-8.

The 12 degrees of freedom corresponding to the *AB* and *ABC* interactions in the factorial experiment appear in Plan 8a as the sum of 6 degrees of freedom for groups within *C* and 6 degrees of freedom for what is

partial information on the simple interaction of AB at each level of c_k . In the general case, the relationship is as follows:

df			df
$(p-1)^2$	AB	(Groups within C	$p(p-1)$
$(p-1)^3$	ABC	$\Sigma(AB'$ for each square)	$\frac{p(p-1)(p-2)}{p(p-1)^2}$
$p(p-1)^2$			

In order to have complete within-subject information on all effects, except the main effects of factor C , p^2 observations on each subject are required. In Plan 8a (constructed with a balanced set of squares) only p

Table 10.7-9 Summary of Computational Procedures for Plan 8

(i)	(1) = G^2/np^2q	(6) = $[\Sigma(AC_{ik})^2]/np$
	(2) = ΣX^2	(7) = $[\Sigma(BC_{jk})^2]/np$
	(3) = $(\Sigma A_i^2)/npq$	(8) = $\Sigma(P_{o(m)}^2)/p$
	(4) = $(\Sigma B_j^2)/npq$	(9) = $[\Sigma(\text{cell total})^2]/n$
	(5) = $(\Sigma C_k^2)/np^2$	(10) = $(\Sigma G_m^2)/npq$
(In each case the summation is over the whole experiment.)		
(ii)	$SS_{\text{between subjects}}$	= (8) - (1)
	SS_c	= (5) - (1)
	$SS_{\text{groups w. } C}$	= (10) - (5)
	$SS_{\text{subj w. groups}}$	= (8) - (10)
	$SS_{\text{w. subj}}$	= (2) - (8)
	SS_a	= (3) - (1)
	SS_b	= (4) - (1)
	SS_{ac}	= (6) - (3) - (5) + (1)
	SS_{bc}	= (7) - (4) - (5) + (1)
	SS_{res}	= (9) - (10) - (6) - (7) + 2(5)
	$SS_{\text{error (within)}}$	= (2) - (9) - (8) + (10)

observations are required on each subject. What is lost in reducing the number of observations per subject from p^2 to p is partial information on the AB and ABC interactions, specifically $p(p-1)$ of these degrees of freedom. This loss in information is, in many cases, a small price to pay for the potential saving in experimental feasibility and effort.

In Plan 8a, the source of variation SS'_{ab} , which is actually part of the simple AB interaction for each of the levels of factor C , may be tested separately or as a sum. As a sum, this source of variation is a mixture of the AB and ABC interactions.

Computational Procedures for Plan 8. Computational procedures for this plan are summarized in Table 10.7-9. These same procedures may also be followed for Plan 8a. In the latter plan, SS_{res} becomes $\Sigma SS'_{ab}$ for each square.

In part i of this table symbols convenient for use in the computational formulas are defined. Computational formulas are given in part ii. The order in which the sums of squares are given in part ii corresponds to the order in Table 10.7-6. For purposes of checking homogeneity of the parts of $SS_{\text{error(within)}}$, this source of variation may be partitioned into pq non-overlapping parts—one part for each of the pq groups in the experiment. The part corresponding to G_m is

$$SS_{\text{error(within) for } G_m} = (2g_m) - (9g_m) - (8g_m) + (10g_m),$$

where $(2g_m)$ designates the numerical value of (2) if the summation in the latter is restricted to G_m . Analogous definitions hold for the other computational symbols containing g_m . It is readily shown that

$$SS_{\text{error(within)}} = \sum SS_{\text{error(within) for } G_m}.$$

The F_{max} statistic may be used to check on the homogeneity of the parts that are pooled to provide the over-all error term for the within-subject effects. It is readily verified that

$$\Sigma(2g_m) = (2), \quad \Sigma(9g_m) = (9), \quad \text{etc.}$$

The variation due to subjects within groups may also be checked for homogeneity. The sum of squares $SS_{\text{subj w. groups}}$ may be partitioned into pq parts, one part for each group. Each part has the general form

$$SS_{\text{subj w. } G_m} = (8g_m) - (10g_m),$$

where $(8g_m)$ is the numerical value of (8) if the summation is restricted to group m .

If a balanced set of squares is used in the construction of Plan 8a, the computational procedures may be simplified. One may use the regular computational procedures for a three-factor factorial experiment having n observations per cell as the starting point. In addition one computes $SS_{\text{groups w. } G}$ and $SS_{\text{subj w. groups}}$ by means of the procedures given for Plan 8. Then

$$\Sigma(SS'_{ab \text{ for each square}}) = SS_{ab} + SS_{abc} - SS_{\text{groups w. } G},$$

$$SS_{\text{error(within)}} = SS_{\text{w. cell}} - SS_{\text{subj w. groups}}.$$

Illustrative Application. Conrad (1958) reports an experiment that may be considered a special case of Plan 8. The purpose of this experiment was to investigate the accuracy with which subjects performed under four methods for dialing telephone numbers. There were 24 subjects in the experiment. Rather than assigning groups of 6 subjects to the rows of a single 4×4 square, individual subjects were assigned at random to six different 4×4 squares. The squares thus represented different sets of

sequences under which the subjects used the methods. The analysis of variance for the Conrad data was as follows:

	df	MS	F
Between subjects	23		
C Squares	5	2,446	
Subj w. squares	18	28,651	
Within subjects	72		
A Methods	3	3,018	9.61
B Order	3	1,106	3.52
AC	15	197	
BC	15	215	
Residual	36	314	

Plan 9. This plan may be viewed as a special case of Plan 8a in which the same square is used for all levels of factor *C*. Hence the same components of the *AB* interaction are confounded within each of the squares, and the same components of the *ABC* interaction are confounded in differences between squares. Full information will be available on some components of both the *AB* and *ABC* interactions. The number of levels of factors *A* and *B* must be equal; there is no restriction on the number of levels of factor *C*. Factors *A*, *B*, and *C* are considered to be fixed; interactions with the group factor are assumed to be negligible. A schematic representation of this plan for the case in which $p = 3$ and $q = 3$ is given below:

I					II					III				
		a_1	a_2	a_3			a_1	a_2	a_3			a_1	a_2	a_3
c_1	G_1	b_2	b_3	b_1	c_2	G_4	b_2	b_3	b_1	c_3	G_7	b_2	b_3	b_1
	G_2	b_1	b_2	b_3		G_5	b_1	b_2	b_3		G_8	b_1	b_2	b_3
	G_3	b_3	b_1	b_2		G_6	b_3	b_1	b_2		G_9	b_3	b_1	b_2

It is of considerable interest to indicate the manner in which the *AB* and *ABC* interactions are confounded. A given row in each of the above squares represents the same combination of treatments *A* and *B*. For example, the first row in each square involves ab_{12} , ab_{23} , and ab_{31} . A summary table of the following form will be shown to have special meaning in the interpretation of the *AB* and *ABC* interactions:

Row	c_1	c_2	c_3	Total
1	G_1	G_4	G_7	R_1
2	G_2	G_5	G_8	R_2
3	G_3	G_6	G_9	R_3
	C_1	C_2	C_3	Grand total

In this summary table each G designates the sum of the np observations made within a group.

The total R_1 represents the sum of the npq observations in row 1 of all squares; this sum is also $G_1 + G_4 + G_7$. This total is balanced with respect to the main effects of factors A , B , and C ; it is also balanced with respect to the ABC interaction. The latter balance follows from the fact that sums of the form

$$\alpha\beta\gamma_{121} + \alpha\beta\gamma_{122} + \alpha\beta\gamma_{123} = 0,$$

$$\alpha\beta\gamma_{231} + \alpha\beta\gamma_{232} + \alpha\beta\gamma_{233} = 0.$$

However, the total R_1 is not balanced with respect to the AB interaction, since

$$\alpha\beta_{12} + \alpha\beta_{23} + \alpha\beta_{31} \neq 0.$$

By similar reasoning each of the other R 's may be shown to be balanced with respect to all main effects as well as the ABC interaction; however, each of the other R 's is not balanced with respect to the AB interaction. Thus, variation due to differences between the R 's represents two of the four degrees of freedom of the AB interaction. In general such differences represent $p - 1$ of the $(p - 1)^2$ degrees of freedom of the AB interaction.

Since differences between the rows define $p - 1$ degrees of freedom of the AB interaction, the row $\times C$ interaction will define $(p - 1)(q - 1)$ degrees of freedom of the $AB \times C$ interaction. The latter interaction is equivalent to the ABC interaction. Hence $(p - 1)(q - 1)$ of the $(p - 1)(p - 1)(q - 1)$ degrees of freedom for the ABC interaction are confounded with row effects. The remaining degrees of freedom are not confounded with row effects; that is, $(p - 1)(p - 2)(q - 1)$ degrees of freedom of the ABC interaction are within-subject effects.

In some texts the following notation has been used.

$$SS_{\text{rows}} = SS_{ab(\text{between})},$$

$$SS_{\text{row} \times C} = SS_{abc(\text{between})}.$$

In terms of this notation,

$$SS_{ab} = SS_{ab(\text{between})} + SS_{ab(\text{within})}.$$

The corresponding degrees of freedom are

$$(p - 1)(p - 1) = (p - 1) + (p - 1)(p - 2).$$

The three-factor interaction takes the following form:

$$SS_{abc} = SS_{abc(\text{between})} + SS_{abc(\text{within})},$$

$$(p - 1)(p - 1)(q - 1) = (p - 1)(q - 1) + (p - 1)(p - 2)(q - 1).$$

In each case the "between" component is part of the between-subject variation, and the "within" component is part of the within-subject effects.

The model under which the analysis of variance for this plan may be carried out is

$$E(X_{ijkmo}) = \mu + \gamma_k + (\text{row})_m + (\gamma \times \text{row})_{km} + \pi_{o(m)} + \alpha_i + \beta_j + \alpha\beta'_{ij} \\ + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma'_{ijk}.$$

The analysis corresponding to this model is summarized in Table 10.7-10. The expected values of the mean squares in this table are derived under the assumptions that factors A , B , and C are fixed; the group factor and subjects within groups are considered random. All interactions with the group and subject effects are considered negligible.

Table 10.7-10 Analysis of Plan 9

Source of variation	df	df for general case	E(MS)
Between subjects	$9n - 1$	$npq - 1$	
C	2	$q - 1$	$\sigma_\epsilon^2 + p\sigma_\pi^2 + np^2\sigma_\gamma^2$
Rows [AB (between)]	2	$p - 1$	$\sigma_\epsilon^2 + p\sigma_\pi^2 + nq\sigma_{\alpha\beta}^2$
$C \times \text{row}$ [ABC (between)]	4	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + p\sigma_\pi^2 + n\sigma_{\alpha\beta}^2$
Subjects w. groups	$9(n - 1)$	$pq(n - 1)$	$\sigma_\epsilon^2 + p\sigma_\pi^2$
Within subjects	$18n$	$npq(p - 1)$	
A	2	$p - 1$	$\sigma_\epsilon^2 + npq\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_\epsilon^2 + npq\sigma_\beta^2$
AC	4	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + np\sigma_{\alpha\gamma}^2$
BC	4	$(p - 1)(q - 1)$	$\sigma_\epsilon^2 + np\sigma_{\beta\gamma}^2$
$(AB)'$	2	$(p - 1)(p - 2)$	$\sigma_\epsilon^2 + nq\sigma_{\alpha\beta}^2$
$(ABC)'$	4	$(p - 1)(p - 2)(q - 1)$	$\sigma_\epsilon^2 + n\sigma_{\alpha\beta\gamma}^2$
Error (within)	$18n - 18$	$pq(p - 1)(n - 1)$	σ_ϵ^2

The analysis given in this table indicates that Plan 9 may be considered as a special case of a $p \times p \times q$ factorial experiment arranged in blocks of size p , the groups having the role of blocks. The differences between the groups may be partitioned as follows:

Between groups	$pq - 1$
C	$q - 1$
Groups within C	$q(p - 1)$
Rows	$p - 1$
$C \times \text{rows}$	$(p - 1)(q - 1)$

What corresponds to the within-cell variation in a $p \times p \times q$ factorial experiment having n observations per cell is partitioned as follows in Plan 9.

Within cell	$p^2q(n - 1)$
Subjects w. groups	$pq(n - 1)$
Error (within)	$pq(n - 1)(p - 1)$

Numerical Example of Plan 9. Suppose that a research worker is interested in evaluating the relative effectiveness of three variables in the design of a package. The variables to be studied are kind of material, style of printing to be used, and color of the material. Each of these variables constitutes a factor; there are three levels under each factor. That is, there are three kinds of material, three styles of printing, and three colors. A total of $3 \times 3 \times 3 = 27$ different packages can be constructed. Suppose that all 27 packages are constructed.

If the research worker selects Plan 9 for use, each subject is required to judge only 3 of the 27 packages. Under the conditions of the study, this number is considered to be the largest number feasible for any one subject. One of the main effects in Plan 9 (that of factor C) is wholly a between-subject effect, but within-subject information is available with respect to the interactions with that factor. Anticipating interactions with the color factor, and desiring within-subject information on interactions with the color factor, the research worker decides to let color correspond to factor C in Plan 9. This plan is symmetrical with respect to factors A and B ; hence there is no difference whether the material factor corresponds to factor A or B . Suppose that the material factor is made to correspond to factor A .

The number of subjects to be used in each of the groups depends upon the precision that is desired. In related studies of this kind a minimum of 10 to 20 subjects per group is generally required to obtain a satisfactory power. Depending upon the degree of variability of the judgments, a larger or smaller number of subjects will be required. For purposes of illustrating the computational procedures, suppose that there are only two subjects in each group. The basic observational data for this illustration are given in Table 10.7-11. The design is that given earlier in this section; the data have been rearranged for computational convenience.

The subjects are asked to rate the packages assigned to them on a 15-point criterion scale. Subjects are assigned at random to the groups; the order in which a subject judges a set of packages is randomized independently for each subject. Subjects in groups 1, 4, and 7 each judge combinations ab_{12} , ab_{23} , and ab_{31} . However, these combinations appear in a different color (factor C) for each of the three groups. Person 1 made the ratings 2, 2, and 3 for the respective combinations, all of which are at color c_1 . Person 6 made the ratings 7, 5, and 9 for the respective combinations—these combinations are in color c_3 . The assignment of the treatment combinations to the groups follows the schematic representation of Plan 9 given earlier in this section. For ease in computation the data in Table 10.7-11 have been rearranged. In addition to the basic observed data, totals needed in the analysis of variance are given at the right.

Summary data obtained from Table 10.7-11 appear in parts i and ii of Table 10.7-12. An ABC summary table appears in part i. Entries in part i

Table 10.7-11 Numerical Example of Plan 9

	Group	Person	a_1	a_2	a_3	Person total	Group total	Row total
			b_2	b_3	b_1			
c_1	G_1	1 2	2 1	2 1	3 7	7 = P_1 9	16 = G_1	
c_2	G_4	3 4	5 9	8 12	1 7	14 28		42
c_3	G_7	5 6	5 7	4 5	6 9	15 21	36	94 = R_1
			b_1	b_2	b_3			
c_1	G_2	7 8	5 8	4 5	7 8	16 21	37	
c_2	G_5	9 10	8 10	10 14	4 6	22 30		52
c_3	G_8	11 12	10 7	10 3	8 9	28 19	47	136
			b_3	b_1	b_2			
c_1	G_3	13 14	3 6	2 4	5 9	10 19	29	
c_2	G_6	15 16	8 10	9 10	6 5	23 25		48
c_3	G_9	17 18	12 8	6 8	10 2	28 18	46	123
Total			124	117	112	353 = G	353	353

have the general symbol ABC_{ijk} and are each based upon two observations. For example, the entry ABC_{111} is the sum of all observations made under treatment combination abc_{111} . Only subjects 7 and 8 are observed under this treatment combination. Hence

$$ABC_{111} = 5 + 8 = 13.$$

As another example, only subjects 11 and 12 are observed under treatment combination abc_{333} ; hence

$$ABC_{333} = 8 + 9 = 17.$$

Table 10.7-12 Summary Data for Numerical Example

	c_1			c_2			c_3			Total
	a_1	a_2	a_3	a_1	a_2	a_3	a_1	a_2	a_3	
b_1	13	6	10	18	19	8	17	14	15	120
b_2	3	9	14	14	24	11	12	13	12	112
b_3	9	3	15	18	20	10	20	9	17	121
										353

	a_1	a_2	a_3	Total		a_1	a_2	a_3	Total
b_1	48	39	33	120	c_1	25	18	39	82
b_2	29	46	37	112	c_2	50	63	29	142
b_3	47	32	42	121	c_3	49	36	44	129
Total	124	117	112	353	Total	124	117	112	353

	b_1	b_2	b_3	Total		Row 1	Row 2	Row 3	Group total
c_1	29	26	27	82	c_1	16	37	29	82
c_2	45	49	48	142	c_2	42	52	48	142
c_3	46	37	46	129	c_3	36	47	46	129
Total	120	112	121	353		94	136	123	353

(iii)	(1)	$= G^2/np^2q$	$= 2307.57$
	(2)	$= \Sigma X^2$	$= 2811$
	(3)	$= (\Sigma A_i^2)/npq$	$= 2311.61$
	(4)	$= (\Sigma B_j^2)/npq$	$= 2310.28$
	(5)	$= (\Sigma C_k^2)/np^2$	$= 2418.28$
	(6)	$= [\Sigma (AB_{ij})^2]/nq$	$= 2372.83$
	(7)	$= [\Sigma (AC_{ik})^2]/np$	$= 2568.83$
	(8)	$= [\Sigma (BC_{jk})^2]/np$	$= 2429.50$
	(9)	$= [\Sigma (ABC_{ijk})^2]/n$	$= 2654.50$
	(10)	$= (\Sigma P_o^2)/p$	$= 2575.00$
	(11)	$= (\Sigma G_m^2)/np$	$= 2476.50$
	(12)	$= (\Sigma R_s^2)/npq$	$= 2358.94$

From the data in part i, AB , AC , and BC summary tables may be prepared; these appear in part ii. Data for the $C \times$ row summary table in part ii are obtained from the column headed Group total in Table 10.7-11. Computational symbols convenient for use in obtaining the required sums of squares are defined in part iii. The numerical values of these symbols for the data are also given.

The analysis of variance is summarized in Table 10.7-13. Factors A , B , and C are considered to be fixed. Tests on the between-subject effects

Table 10.7-13 Analysis of Variance for Numerical Example

Source of variation	Computational formula	SS	df	MS	F
Between subjects		267.43	17		
<i>C</i> (color)	(10) - (1)	110.71	2	55.36	5.06*
Rows	(5) - (1)	51.37	2	25.68	2.35
<i>C</i> × row	(12) - (1)	6.85	4	1.71	
Subjects within group	(11) - (5) - (12) + (1)	98.50	9	10.94	
Within subjects	(10) - (11)				
	(2) - (10)	236.00	36		
<i>A</i> (material)	(3) - (1)	4.04	2	2.02	
<i>B</i> (printing)	(4) - (1)	2.71	2	1.36	
<i>AC</i>	(7) - (3) - (5) + (1)	146.51	4	36.63	
<i>BC</i>	(8) - (4) - (5) + (1)	8.51	4	2.13	
<i>AB'</i>	[(6) - (3) - (4) + (1)] - [(12) - (1)]	7.14	2	3.57	
<i>ABC'</i>	[(9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)]				
	- [(11) - (5) - (12) + (1)]	9.09	4	2.27	
Error (within)	(2) - (10) - (9) + (11)	58.00	18	3.22	
					11.38**

use $MS_{\text{subj w. groups}}$ in the denominator of F ratios. Tests on within-subject effects use $MS_{\text{error(within)}}$. It will be noted that the denominator for the within-subject effects (3.22) is considerably smaller than the denominator for the between-subject effects (10.94).

Table 10.7-14 Analysis of Simple Effects for Factor A

Source of variation	Computational formula	SS	df	MS	F
A at level c_1	$(3c_1) - (1c_1)$	38.11	2	19.06	5.92*
A at level c_2	$(3c_2) - (1c_2)$	98.11	2	49.06	15.24**
A at level c_3	$(3c_3) - (1c_3)$	14.33	2	7.16	2.22
Error (within)	(From Table 12.7-13)	150.55	18	3.22	

The F tests indicate a highly significant AC interaction. To help interpret this interaction effect, the profiles for simple main effects of factor A (materials) at the various levels of factor C (colors) are given in Fig. 10.1. These profiles were prepared from the AC summary table in part ii of Table 10.7-12. Although color c_2 has the highest over-all average rating, c_2 in combination with material a_3 has one of the lower average ratings. Material a_2 in combination with color c_2 has the highest average rating. An analysis of the simple main effects of factor A for the levels of factor C will permit statistical tests of the differences among the set of points in the same profile. These tests are summarized in Table 10.7-14. Differences between the points on the profile for c_2 are statistically significant ($F = 15.24$). A test on the difference between a_1 and a_2 at level c_2 is given by

$$\begin{aligned}
 F &= \frac{(AC_{12} - AC_{22})^2}{2np \text{ MS}_{\text{error(within)}}} \\
 &= \frac{(50 - 63)^2}{2(6)(3.22)} = 4.37.
 \end{aligned}$$

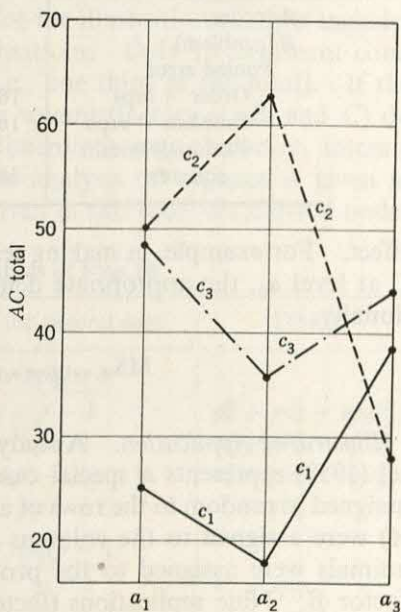


Figure 10.1

For a .05-level test, the critical value for this last test is $F_{.95}(1,18) = 4.41$. Hence the experimental data indicate a difference in favor of material a_2 , but this difference is not quite large enough to be considered statistically significant. If the costs of materials a_1 and a_2 are equal, the experimental

data indicate that a_2 is to be preferred to a_1 when color c_2 is used. However, if there are differential costs, the decision must take such cost differentials into consideration.

Tests on simple main effects are generally made along the dimension of greatest interest to the experimenter. This dimension, by design, is most frequently the within-subject dimension. However, should the experimenter desire to make tests on the simple main effects of what corresponds to factor C , the denominator of such tests is not $MS_{\text{subj w. groups}}$. Rather the denominator is $MS_{\text{w. cell}}$ for the appropriate level of the simple main

Table 10.7-15 Analysis of Variance for Michels et al. Experiment

Source	df	MS	<i>F</i>
Between subjects	26		
Sequences (rows)	2	3.05	
<i>C</i> (replications)	8	5.71	
Sequence \times reps	16	7.59	
Within subjects	54		
<i>A</i> (order)	2	56.64	5.12
<i>B</i> (problem)	2	417.64	37.73
Pooled error	50	11.07	
Order \times reps	16	12.89	
Problem \times reps	16	9.43	
Residuals within squares	18	10.90	

effect. For example, in making tests on the simple main effects of factor C at level a_2 , the appropriate denominator is $MS_{\text{w. cell (for } a_2 \text{)}}$. Computationally,

$$MS_{\text{w. cell (for } a_2 \text{)}} = \frac{(2a_2) - (9a_2)}{pq(n-1)}.$$

Illustrative Application. A study reported by Michels, Bevan, and Straszel (1958) represents a special case of Plan 9. Individual animals were assigned at random to the rows of a 3×3 Latin square. Problems (factor A) were assigned to the columns of the square; the order in which the animals were assigned to the problems was determined by the levels of factor B . Nine replications (factor C) of the experiment were run; the same Latin square was used throughout. The criterion was the number of trials required to learn the problems.

Since the same square was used in each of the replications, corresponding rows of the squares represent a sequence effect. The experiment was actually conducted in distinct replications. Factor C , the replication factor, was considered to be a random factor. The analysis of variance

reported by the authors is summarized in Table 10.7-15. Since the interactions with the replication factor were homogeneous and did not differ from the residuals within the squares, a single pooled error term was used in making all within-subject tests.

Plan 10. This plan is essentially an extension of Plan 6; more explicitly, Plan 6 is the building block from which Plan 10 is constructed. The latter is a series of such building blocks, each at a different level of factor D . Alternatively, Plan 10 may be viewed as a fractional replication of a $p \times p \times p \times q$ factorial experiment. As an illustration, a $3 \times 3 \times 3 \times 2$ factorial experiment will be used. A schematic representation of this plan is given below:

			a_1	a_2	a_3				a_1	a_2	a_3
d_1	G_1	c_1	b_3	b_1	b_2	d_2	G_4	c_1	b_2	b_1	b_3
	G_2	c_2	b_2	b_3	b_1		G_5	c_2	b_1	b_3	b_2
	G_3	c_3	b_1	b_2	b_3		G_6	c_3	b_3	b_2	b_1

As shown above, a different square is used for the two levels of factor D . In cases where partial information on confounded interactions is to be recovered, it is desirable to use the same square for each level of factor D .

The complete factorial experiment for the illustrative example includes $3 \times 3 \times 3 \times 2 = 54$ treatment combinations. Only 18 treatment combinations appear in the above plan (i.e., one third of the total). If the factors forming the dimensions of the square (factors A , B , and C) do not interact with each other (these dimensions may, however, interact with factor D), then an outline of the analysis of variance is given in Table 10.7-16. The expected values given in this table are derived under

Table 10.7-16 Analysis of Plan 10

Source of variation	df	df for general case	E(MS)
Between subjects	$6n - 1$	$npq - 1$	
$C (AB')$	2	$p - 1$	$\sigma_e^2 + p\sigma_\pi^2 + npq\sigma_\gamma^2$
D	1	$q - 1$	$\sigma_e^2 + p\sigma_\pi^2 + np^2\sigma_\delta^2$
$CD (AB' \times D)$	2	$(p - 1)(q - 1)$	$\sigma_e^2 + p\sigma_\pi^2 + np\sigma_{\gamma\delta}^2$
Subjects within groups	$6(n - 1)$	$pq(n - 1)$	$\sigma_e^2 + p\sigma_\pi^2$
Within subjects	$12n$	$npq(p - 1)$	
A	2	$p - 1$	$\sigma_e^2 + npq\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_e^2 + npq\sigma_\beta^2$
AD	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\alpha\delta}^2$
BD	2	$(p - 1)(q - 1)$	$\sigma_e^2 + np\sigma_{\beta\delta}^2$
$(AB)''$	2	$(p - 1)(p - 2)$	$\sigma_e^2 + nq\sigma_{\alpha\beta}^2$
$(AB)'' \times D$		$(p - 1)(p - 2)(q - 1)$	$\sigma_e^2 + n\sigma_{\alpha\beta\delta}^2$
Error (within)	$12n - 12$	$pq(n - 1)(p - 1)$	σ_e^2

the assumption that factors A , B , C , and D are fixed. Groups and subjects within groups are considered random.

If, for example, the AB interaction were not negligible, then the main effects of factor C and the CD interaction would be confounded with this interaction. A partial check on the assumption that dimensions of the Latin square do not interact with each other is provided by the F ratio

$$F = \frac{MS_{\text{res}}}{MS_{\text{error(within)}}}$$

A somewhat better check on these assumptions is obtained by computing the residual and error (within) separately for each level of factor D as well as the corresponding pooled terms. Individual tests are made for the separate levels as well as for the pooled terms.

Relative to what would be the case in a complete factorial experiment, the within-cell variation is partitioned as follows:

Within cell	$p^2q(n-1)$
Subjects w. groups	$pq(n-1)$
Error (within)	$pq(n-1)(p-1)$

All other sources of variation are part of the between-cell variation; the degrees of freedom for the latter are $p^2q - 1$.

Plan 10 is useful in situations illustrated by the following example: Suppose that a research worker is interested in evaluating the effectiveness of q different methods of training (factor D) on marksmanship. There are p different kinds of targets (factor B) to be used in the evaluation process. Each target is to be placed at p different distances (factor C). Subjects are to fire at each of the p targets, but subjects are assigned to only one distance. The order (factor A) in which subjects fire at the targets is balanced by means of a Latin square. In this kind of experiment, primary interest is generally in the main effects of factor D and the interactions with factor D , particularly the BD and CD interactions. The BD interaction is a within-subject effect; the CD interaction, a between-subject effect. If the experimenter has the choice of what variables are assigned as factors B and C , factor B should be the one on which the more precise information is desired. The experimenter may not always have his choice in such matters—the dictates of experimental feasibility frequently force a decision of this kind in a given direction.

Computational Procedures. Computational procedures for this plan are summarized in Table 10.7-17. Since all treatment combinations in the factorial experiment involving factors C and D appear in this plan, a CD summary table may be prepared. From the latter summary table one may compute sums of squares of the main effects of C and D as well as the sum of squares for their interaction. One may also prepare AD and BD

summary tables, since all the treatment combinations in corresponding two-factor factorial experiments occur for these two sets of variables.

For the basic observations one may obtain each of the P_o , where the latter symbol denotes the sum of the p observations on an individual subject. The expression (cell total) used in computational symbol (11) denotes the sum of the n observations made under a specified treatment combination. Another symbol for this total is $ABCD_{ijkm}$.

Table 10.7-17 Computational Procedures for Plan 10

(i)	(1) = G^2/np^2q	(7) = $[\Sigma(AD_{im})^2]/np$
	(2) = ΣX^2	(8) = $[\Sigma(BD_{jm})^2]/np$
	(3) = $(\Sigma A_i^2)/npq$	(9) = $[\Sigma(CD_{km})^2]/np$
	(4) = $(\Sigma B_j^2)/npq$	(10) = $(\Sigma P_o^2)/p$
	(5) = $(\Sigma C_k^2)/npq$	(11) = $[\Sigma(\text{cell total})^2]/n$
	(6) = $(\Sigma D_m^2)/np^2$	
<hr/>		
(ii)	Between subjects	(10) - (1)
	C	(5) - (1)
	D	(6) - (1)
	CD	(9) - (5) - (6) + (1)
	Subjects within groups	(10) - (9)
	Within subjects	(2) - (10)
	A	(3) - (1)
	B	(4) - (1)
	AD	(7) - (3) - (6) + (1)
	BD	(8) - (4) - (6) + (1)
<hr/>		
	Residual	(11) - (7) - (8) - (9) + 2(6)
	Error (within)	(2) - (10) - (11) + (9)

The residual variation is that part of the variation between cells remaining after (presumably) all the unconfounded sources of variation due to treatment effects have been taken into account. In symbols,

$$SS_{\text{res}} = SS_{\text{between cells}} - (SS_a + SS_b + SS_c + SS_d + SS_{cd} + SS_{ad} + SS_{bd}).$$

The homogeneity of the residual variation may be checked by computing separate residual terms for each Latin square. Since each of the Latin squares represents a different level of factor D , the latter procedure is equivalent to computing a separate residual term for each level of factor D .

$$SS_{\text{res for level } d_m} = (11d_m) - (7d_m) - (8d_m) - (9d_m) + 2(6d_m),$$

where the symbol $(11d_m)$ is (11) with the summation restricted to level d_m .

The variation due to error (within) is that part of the within-cell variation which remains after variation due to differences between subjects within cells is removed. In symbols,

$$SS_{\text{error(within)}} = SS_{\text{w. cell}} - SS_{\text{subj w. groups}}.$$

The latter term may be checked for homogeneity by computing a separate $SS_{\text{error(within)}}$ for each group. Computationally,

$$SS_{\text{error(within)}} \text{ for } G_r = (2g_r) - (10g_r) - (11g_r) + (9g_r).$$

Plan 11. In Plan 7 only a fraction of the treatment combinations in a $p \times p \times p$ factorial experiment actually appears in the experiment. However, in Plan 11 all the treatment combinations in this factorial experiment are used. The primary purpose of this plan is to obtain complete within-subject information on all main effects, as well as partial within-subject information on all interaction effects in a $p \times p \times p$ factorial experiment. Yet only p observations on each subject are required.

Table 10.7-18 Schematic Representation of Plan 11

Schematic Representation of Plan II												
$a_1 \quad a_2 \quad a_3$				$a_1 \quad a_2 \quad a_3$				$a_1 \quad a_2 \quad a_3$				
(i)	G_1	bc_{12}	bc_{31}	bc_{23}	G_4	bc_{32}	bc_{21}	bc_{13}	G_7	bc_{22}	bc_{11}	bc_{33}
	G_2	bc_{33}	bc_{22}	bc_{11}	G_5	bc_{23}	bc_{12}	bc_{31}	G_8	bc_{13}	bc_{32}	bc_{21}
	G_3	bc_{21}	bc_{13}	bc_{32}	G_6	bc_{11}	bc_{33}	bc_{22}	G_9	bc_{31}	bc_{23}	bc_{12}
(ii)	G_1	(112)	(231)	(323)	G_4	(132)	(221)	(313)	G_7	(122)	(211)	(333)
	G_2	(133)	(222)	(311)	G_5	(123)	(212)	(331)	G_8	(113)	(232)	(321)
	G_3	(121)	(213)	(332)	G_6	(111)	(233)	(322)	G_9	(131)	(223)	(312)

The construction of this plan will be illustrated by means of a $3 \times 3 \times 3$ factorial experiment. The starting point for this plan is a set of balanced 3×3 Latin squares.

I			II			III		
1	3	2	3	2	1	2	1	3
3	2	1	2	1	3	1	3	2
2	1	3	1	3	2	3	2	1

It is noted that square II is obtained from square I by means of a one-step cyclic permutation of the rows. Similarly, square III is obtained from square II by a one-step cyclic permutation. One now constructs a square orthogonal to square I. The following square has this property:

I'		
2	1	3
3	2	1
1	3	2

This square is also orthogonal to squares II and III.

In the design given in part i of Table 10.7-18, the subscripts for factor B are determined by the corresponding numbers in squares I, II, and III. The subscripts for factor C are determined by the numbers in square I'. A different notation system is used in part ii of this table—the numbers in parentheses represent the respective subscripts for the treatment combinations.

The n subjects in each group are observed under the treatment combinations in each of the rows. These sets are balanced with respect to main effects but only partially balanced with respect to interaction effects. For purposes of illustrating the manner in which the AB interaction is partially confounded with differences between groups, consider the sets of treatment combinations assigned to the following groups:

G_1	(112)	(231)	(323)
G_6	(111)	(233)	(322)
G_8	(113)	(232)	(321)

Within each of these groups there are repeated measures on the same set of combinations of factor A and B , namely, ab_{11} , ab_{23} , and ab_{32} . There is balance with respect to main effects as well as the AC , BC , and ABC interactions when the sum of all observations in the three groups is obtained. If all factors are fixed, it may be shown by direct substitution in the basic linear model that

$$E(G_{1+6+8}) = 9n\mu + 3n(\alpha\beta_{11} + \alpha\beta_{23} + \alpha\beta_{32}) + 3n(\delta_1 + \delta_6 + \delta_8).$$

The symbol G_{1+6+8} denotes the sum of the $9n$ observations in these three groups. The effects δ_1 , δ_6 , and δ_8 designate group effects. Since G_2 , G_4 , and G_8 are each observed under the set of treatment combinations ab_{13} , ab_{22} , and ab_{31} , and since there is balance with respect to the main effects as well as the AC and BC interactions,

$$E(G_{2+4+9}) = 9n\mu + 3n(\alpha\beta_{13} + \alpha\beta_{22} + \alpha\beta_{31}) + 3n(\delta_2 + \delta_4 + \delta_9).$$

It may also be shown that

$$E(G_{3+5+7}) = 9n\mu + 3n(\alpha\beta_{12} + \alpha\beta_{21} + \alpha\beta_{33}) + 3n(\delta_3 + \delta_5 + \delta_7).$$

Thus differences between the three totals G_{1+6+8} , G_{2+4+9} , and G_{3+5+7} are in part due to the AB interaction and in part due to group effects. Hence two of the four degrees of freedom of the AB interaction are partially confounded with differences between groups.

In Plan 11, as constructed in Table 10.7-18, two-factor interactions are partially confounded with the sets of group totals given below:

Interaction	Sets of group totals
AB	G_{1+6+8} , G_{2+4+9} , G_{3+5+7}
AC	G_{1+4+7} , G_{2+5+8} , G_{3+6+9}
BC	G_{1+5+9} , G_{2+6+7} , G_{3+4+8}

In each case two of the four degrees of freedom for the respective two-factor interactions are confounded with differences between groups. The remaining degrees of freedom for the variation due to differences between groups are confounded with the three-factor interaction. The totals involved in the latter are G_{1+2+3} , G_{4+5+6} , and G_{7+8+9} . An outline of the analysis of variance appears in Table 10.7-19.

Table 10.7-19 Analysis of Plan 11

Source of variation	df	df for general case	E(MS)
Between subjects	$9n - 1$	$np^2 - 1$	
Groups	8	$p^2 - 1$	
(AB)'	2	$p - 1$	
(AC)'	2	$p - 1$	
(BC)'	2	$p - 1$	
(ABC)'	2	$(p - 1)(p - 2)$	
Subjects within groups	$9n - 9$	$p^2(n - 1)$	
Within subjects	$18n$	$np^2(p - 1)$	
A	2	$p - 1$	$\sigma_e^2 + np^2\sigma_\alpha^2$
B	2	$p - 1$	$\sigma_e^2 + np^2\sigma_\beta^2$
C	2	$p - 1$	$\sigma_e^2 + np^2\sigma_\gamma^2$
(AB)''	2	$(p - 1)(p - 2)$	$\sigma_e^2 + np\sigma_{\alpha\beta}^2$
(AC)''	2	$(p - 1)(p - 2)$	$\sigma_e^2 + np\sigma_{\alpha\gamma}^2$
(BC)''	2	$(p - 1)(p - 2)$	$\sigma_e^2 + np\sigma_{\beta\gamma}^2$
(ABC)''	6	$(p - 1)^3 - (p - 1)(p - 2)$	$\sigma_e^2 + n\sigma_{\alpha\beta\gamma}^2$
Error (within)	$18n - 18$	$p^2(p - 1)(n - 1)$	σ_e^2

This plan may be improved by constructing a replication in which different components of the interactions are confounded with group effects. This replication may be obtained from a second set of balanced squares. The following balanced set is different from the original set:

I			II			III		
1	2	3	2	3	1	3	1	2
2	3	1	3	1	2	1	2	3
3	1	2	1	2	3	2	3	1

The following square is orthogonal to square I:

I'		
1	2	3
3	1	2
2	3	1

Table 10.7-20 Replication of Plan 11

G_1 (111) (222) (333)	G_4 (121) (232) (313)	G_7 (131) (212) (323)
G_2 (123) (231) (312)	G_5 (133) (211) (322)	G_8 (113) (221) (332)
G_3 (132) (213) (321)	G_6 (112) (223) (331)	G_9 (122) (233) (311)

A plan constructed from this set of squares is given in Table 10.7-20. For this replication, it may be shown that

$$E(G_{1+6+8}) = 9n + 3n(\alpha\beta_{11} + \alpha\beta_{22} + \alpha\beta_{33}) + 3n(\delta_1 + \delta_6 + \delta_8).$$

The set of $\alpha\beta$'s included in this expected value is different from that included

in the original plan. For this replication it may be shown that variation among the totals

$$G_{1+6+8}, \quad G_{2+4+9}, \quad \text{and} \quad G_{3+5+7}$$

is confounded with two of the four degrees of freedom of the AB interaction; however, the two degrees of freedom that are confounded here are not identical with those confounded in the original plan.

In terms of the notation of Chap. 8 on components of interaction, in the original plan the AB^2 components of the $A \times B$ interaction are confounded with group effects; in the replication the AB components are confounded with group effects. Hence the original plan provides within-subject information on the AB^2 components, and the replication provides

Table 10.7-21 Definition of Computational Symbols

	(1) = G^2/np^3	(6) = $[\Sigma(AB_{ij})^2]/np$
	(2) = ΣX^2	(7) = $[\Sigma(AC_{ik})^2]/np$
(i)	(3) = $(\Sigma A_i^2)/np^2$	(8) = $[\Sigma(BC_{jk})^2]/np$
	(4) = $(\Sigma B_j^2)/np^2$	(9) = $[\Sigma(ABC_{ijk})^2]/n$
	(5) = $(\Sigma C_k^2)/np^2$	(10) = $(\Sigma P_o^2)/p$
	(11) = $(\Sigma G_m^2)/np$	
(ii)	(12) = $(\Sigma G_{ab}^2)/np^2 = (G_{1+6+8}^2 + G_{2+4+7}^2 + G_{3+5+7}^2)/np^2$	
	(13) = $(\Sigma G_{ac}^2)/np^2 = (G_{1+4+7}^2 + G_{2+5+8}^2 + G_{3+6+9}^2)/np^2$	
	(14) = $(\Sigma G_{bc}^2)/np^2 = (G_{1+5+9}^2 + G_{2+6+7}^2 + G_{3+4+8}^2)/np^2$	

within-subject information on the AB components. Similarly, in the original plan the AC components of $A \times C$ are confounded with group effects; in the replication the AC^2 components are confounded. A comparable condition holds for the $B \times C$ interaction. Use of the replication will provide some within-subject information on all components of the two-factor interactions. However, additional replications are required to obtain within-subject information on all components of the three-factor interaction. The original plan provides within-subject information on the AB^2C^2 component; the replication provides within-subject information on the ABC component. No within-subject information is available on the ABC^2 or the AB^2C^2 components.

Computational Procedures for Plan 11. Computational procedures here are similar to those used in a $p \times p \times p$ factorial experiment in which there are n observations per cell. The first steps actually duplicate the latter procedures. The nonreplicated version of Plan 11 will be considered first; the replicated plan will be considered later.

Since all the treatment combinations in a $p \times p \times p$ factorial experiment appear in Plan 11, an ABC summary table may be prepared from the basic observations. From the latter summary table one obtains AB , AC , and

BC summary tables. Most of the computational symbols in part i of Table 10.7-21 are obtained from these summary tables. Symbol (10) involves the totals P_o ; the latter are the sums of the set of observations on an individual subject. The computational symbols in part ii require special comment.

Table 10.7-22 Computational Formulas

Between subjects	(10) - (1)
Between groups	(11) - (1)
AB'	(12) - (1)
AC'	(13) - (1)
BC'	(14) - (1)
ABC'	(11) - (12) - (13) - (14) + 2(1)
Subjects within groups	(10) - (11)
Within subjects	(2) - (10)
A	(3) - (1)
B	(4) - (1)
C	(5) - (1)
AB''	[(6) - (3) - (4) + (1)] - [(12) - (1)]
AC''	[(7) - (3) - (5) + (1)] - [(13) - (1)]
BC''	[(8) - (4) - (5) + (1)] - [(14) - (1)]
ABC''	[(9) - (6) - (7) - (8) + (3) + (4) + (5) - (1)] - [(11) - (12) - (13) - (14) + 2(1)]
Error (within)	[(2) - (9)] - [(10) - (11)]

The symbol G_m denotes the sum of the np observations in group m . The symbol G_{ab} is the sum of those G_m 's which are assigned to the same set of ab_{ij} . For assignments made in accordance with the principles given in the last section, in a $3 \times 3 \times 3$ experiment the G_{ab} 's are

$$G_{1+6+8}, \quad G_{2+4+9}, \quad \text{and} \quad G_{3+5+7},$$

where $G_{1+6+8} = G_1 + G_6 + G_8$. In general there will be np^2 observations in each of such totals. The G_{ac} 's are made up of the following totals:

$$G_{1+4+7}, \quad G_{2+5+8}, \quad \text{and} \quad G_{3+6+9}.$$

Each of the latter G_m 's which are combined into a single total is assigned to the same set of ac_{ik} . For example, inspection of Table 10.7-18 indicates that groups 1, 4, and 7 are assigned to the sets ac_{12} , ac_{21} , and ac_{33} . The level of factor B changes for the different groups, but the set of ac_{ik} remains the same.

Computational formulas for the sum of squares are summarized in Table 10.7-22. The parts of $SS_{\text{error(within)}}$ may be checked for homogeneity by computing a separate sum of squares for each of the groups. Thus,

$$SS_{\text{error(within)}} \text{ for } G_m = [(2g_m) - (9g_m)] - [(10g_m) - (11g_m)],$$

where the symbol $(2g_m)$ represents the symbol (2), in which the summation is restricted to G_m .

If Plan 11 is replicated, it is generally wise to carry out separate computations for each replication; those terms which are homogeneous may then be combined. Since different components of the two-factor interactions are estimated in each of the two replications, the two-factor interactions are in a sense nested within the replications. Thus,

$$SS''_{ab} = SS''_{ab \text{ from rep 1}} + SS''_{ab \text{ from rep 2}}.$$

Similar relationships hold for other two-factor interactions as well as the three-factor interaction. The error (within) term for the combined replications is given by

$$SS_{\text{error(within)}} = SS_{\text{error(within) from rep 1}} + SS_{\text{error(within) from rep 2}}.$$

In a sense, the error (within) effects are also nested within each replication.

Main effects, however, are not nested within replications. The latter are computed by obtaining $A \times \text{replication}$, $B \times \text{replication}$, and $C \times \text{replication}$ summary tables from the basic data. From such tables main effects and corresponding interactions may be computed. The latter interactions are pooled with error (within) if the components prove to be homogeneous with error (within).

Plan 12. This plan resembles Plan 5 as well as Plan 8 but yet has features that neither of the latter designs has. A schematic representation of Plan 12 is given below:

	c_1			c_2		
	b_1	b_2	b_3	b_1	b_2	b_3
G_1	a_2	a_1	a_3	a_2	a_1	a_3
G_2	a_3	a_2	a_1	a_3	a_2	a_1
G_3	a_1	a_3	a_2	a_1	a_3	a_2

In general, there will be p levels of factor A and p levels of factor B . The same $p \times p$ Latin square is used at each of the q levels of factor C . Plan 12 may be regarded as a fractional replication of a $p \times p \times q$ factorial experiment. If the interaction with the group factor is negligible, complete within-subject information is available for the main effects of factors A , B , and C . Complete within-subject information is also available on the AC and BC interactions; partial within-subject information is available on the AB and ABC interactions.

The analysis of variance is outlined in Table 10.7-23. In obtaining the expected values of the mean squares it is assumed that factors A , B , and C are fixed; subjects within the groups define a random variable. The terms σ_u^2 , σ_v^2 , and σ_w^2 , appearing in the expected values, represent interactions with subject effects. Because factors A and B are dimensions of a Latin square, the $A \times \text{subjects within group}$ interaction cannot be distinguished from the $B \times \text{subjects within group}$ interaction. However, the $C \times \text{subjects within group}$ interaction can be distinguished from the others, since factor C is not

Table 10.7-23 Analysis of Plan 12

Source	df	E(MS)
Between subjects	$np - 1$	
Groups	$p - 1$	
Subj w. groups	$p(n - 1)$	
Within subjects	$np(pq - 1)$	
A	$p - 1$	$\sigma_e^2 + \sigma_u^2 + npq\sigma_\alpha^2$
B	$p - 1$	$\sigma_e^2 + \sigma_u^2 + npq\sigma_\beta^2$
(AB)'	$(p - 1)(p - 2)$	$\sigma_e^2 + \sigma_u^2 + nq\sigma_{\alpha\beta'}^2$
Residual (1)	$p(n - 1)(p - 1)$	$\sigma_e^2 + \sigma_u^2$
C	$q - 1$	$\sigma_e^2 + \sigma_v^2 + np^2\sigma_\gamma^2$
C \times groups	$(p - 1)(q - 1)$	$\sigma_e^2 + \sigma_v^2 + np\sigma_{\gamma\delta}^2$
Residual (2)	$p(n - 1)(q - 1)$	$\sigma_e^2 + \sigma_v^2$
AC	$(p - 1)(q - 1)$	$\sigma_e^2 + \sigma_w^2 + np\sigma_{\alpha\gamma}^2$
BC	$(p - 1)(q - 1)$	$\sigma_e^2 + \sigma_w^2 + np\sigma_{\beta\gamma}^2$
(AB)'C	$(p - 1)(p - 2)(q - 1)$	$\sigma_e^2 + \sigma_w^2 + n\sigma_{\alpha\beta'\gamma}^2$
Residual (3)	$p(n - 1)(p - 1)(q - 1)$	$\sigma_e^2 + \sigma_w^2$
Pooled error = residual (1) + residual (2) + residual (3)		
$df_{\text{pooled error}} = p(n - 1)(pq - 1)$		

Table 10.7-24 First Stage in the Analysis of Plan 12

Source	df
Between subjects	$np - 1$
Groups	$p - 1$
Subjects w. group	$p(n - 1)$
Within subjects	$np(pq - 1)$
B	$p - 1$
B \times group	$(p - 1)^2$
B \times subj w. group	$p(n - 1)(p - 1)$
C	$(q - 1)$
C \times group	$(p - 1)(q - 1)$
C \times subj w. group	$p(n - 1)(q - 1)$
BC	$(p - 1)(q - 1)$
BC \times group	$(p - 1)^2(q - 1)$
BC \times subj w. group	$p(n - 1)(p - 1)(q - 1)$

one of the dimensions of the Latin square. In general, the three residual terms should be pooled into a single error term if there is neither a priori nor experimental evidence for heterogeneity of these components.

The computational procedures for this plan are simplified if the analysis of variance is carried out in two stages. In the first stage one of the dimensions of the Latin square, say, factor A , is disregarded. Then the plan reduces to a $p \times q$ factorial experiment with repeated measures on both factors. The analysis of variance for the first stage is outlined in Table 10.7-24. The detailed computational procedures given in Sec. 7.3 may be adapted for use to obtain the first stage of the analysis.

In the second stage of the analysis, the presence of factor A as a dimension of the Latin square is taken into account. The $B \times$ group interaction is partitioned as follows:

$$\begin{array}{l} B \times \text{group} \\ A \\ AB' \end{array} \qquad \begin{array}{l} (p-1)^2 \\ p-1 \\ (p-1)(p-2) \end{array}$$

The latter interaction term may be obtained by subtraction or by the relation

$$SS_{ab'} = SS_{ab} - SS_{\text{groups}},$$

where SS_{ab} is obtained from an AB summary table for the combined levels of factor C . The $BC \times$ group interaction is partitioned as follows:

$$\begin{array}{l} BC \times \text{group} \\ AC \\ AB'C \end{array} \qquad \begin{array}{l} (p-1)^2 \\ (p-1)(q-1) \\ (p-1)(p-2)(q-1) \end{array}$$

The latter interaction term may be obtained by subtraction from the following relation,

$$SS_{ab'c} = SS_{abc} - SS_{c \times \text{group}},$$

where SS_{abc} is obtained from an ABC summary table.

The residual terms in Table 10.7-23 are equivalent to the following interactions:

$$SS_{\text{res}(1)} = SS_{b \times \text{subj w. group}},$$

$$SS_{\text{res}(2)} = SS_{c \times \text{subj w. group}},$$

$$SS_{\text{res}(3)} = SS_{bc \times \text{subj w. group}}.$$

If only the pooled error is obtained, the latter is given by

$$SS_{\text{pooled error}} = SS_{\text{w. cell}} - SS_{\text{subj w. group}}.$$

Illustrative Application. An experiment reported by Briggs, Fitts, and Bahrck (1957) represents a special case of Plan 12. In this study the dimensions of a 4×4 Latin square were noise level (factor A) and blocks of trials (factor B). Factor C represented three experimental sessions.

There were three subjects in each of the groups. The groups were assigned at random to the rows of the Latin square. The groups were required to track a visual target under various levels of visual noise. There were three experimental sessions; within each session there were four blocks of trials—each block was under a noise level determined by the letters of a randomly selected Latin square. The same square was used in all sessions. The rows of the Latin square define the sequence in which the noise levels were given. Hence the group factor may be considered as a sequence factor.

Table 10.7-25 Analysis of Variance for Briggs et al. Experiment

Source	df	MS	F
Between subjects	11		
<i>G</i> Groups (sequence)	3	160	
Subj w. group	8	1369	
Within subjects	132		
<i>C</i> Sessions	2	2546	13.35
<i>CG</i>	6	16	
<i>C</i> × subj w. group	16	191	
<i>A</i> Noise level	3	5289	31.31
<i>B</i> Blocks of trials	3	330	1.95
<i>AC</i>	6	293	1.73
<i>BC</i>	6	109	
(<i>AB</i>)' Square uniqueness	6	148	
(<i>AB</i>)' × <i>C</i> (<i>C</i> × square uniq.)	12	95	
Residual	72	169	

The analysis of variance reported by the authors differs slightly from that given in Table 10.7-23. Their analysis is outlined in Table 10.7-25. The residual term in this analysis corresponds to residual (1) + residual (3). The *C* × subjects within group term corresponds to residual (2). No attempt was made to interpret the partial information on the *AB* and *ABC* interactions. Instead, these sources of variation were considered to be functions of the particular Latin square selected for use. The latter interpretation is the preferred one if the sequence (group) factor can be considered as contributing toward these interactions. If there is no reason for considering residual (2) as potentially different from residuals (1) and (3), the experimental data do not rule against pooling the equivalent of residuals (1), (2), and (3) into a single pooled error term to be used in testing all within-subject effects.

Plan 13. This plan resembles Plan 9; in this case, however, a Greco-Latin square replaces the Latin square. A schematic representation of

Plan 13 is given below:

		c_1	c_2	c_3			c_1	c_2	c_3
d_1	G_1	ab_{12}	ab_{31}	ab_{23}	d_2	G_4	ab_{12}	ab_{31}	ab_{23}
	G_2	ab_{33}	ab_{22}	ab_{11}		G_5	ab_{33}	ab_{22}	ab_{11}
	G_3	ab_{21}	ab_{13}	ab_{32}		G_6	ab_{21}	ab_{13}	ab_{32}

The same Greco-Latin square is used for each level of factor D .

The analysis of variance for this plan is most easily understood if it is made in two stages. In the first stage, two of the four dimensions of the Greco-Latin square are disregarded. Suppose that factors A and B are

Table 10.7-26 Analysis of Variance for Plan 13

	Source	df
(i)	Between subjects	$npq - 1$
	Rows	$p - 1$
	D	$q - 1$
	$D \times \text{row}$	$(p - 1)(q - 1)$
	Subjects within groups	$pq(n - 1)$
	Within subjects	$npq(p - 1)$
	C	$p - 1$
	$C \times \text{row}$	$(p - 1)^2$
	CD	$(p - 1)(q - 1)$
	$CD \times \text{row}$	$(p - 1)^2(q - 1)$
(ii)	$C \times \text{subj w. group (Error)}$	$pq(n - 1)(p - 1)$
	$C \times \text{row}$	$(p - 1)^2$
	A	$p - 1$
	B	$p - 1$
	$(AB)'$	$(p - 1)(p - 3)$
	$CD \times \text{row}$	$(p - 1)^2(q - 1)$
	AD	$(p - 1)(q - 1)$
	BD	$(p - 1)(q - 1)$
	$(AB)'D$	$(p - 1)(p - 3)(q - 1)$

disregarded. The resulting plan may be considered as a $q \times p \times p$ factorial experiment with repeated measures on one of the factors. The analysis of variance for the first stage appears in the upper part of Table 10.7-26.

In the second stage, the interactions which involve two dimensions of the square (C and rows) are partitioned into main effects and interactions associated with factors A and B . This stage of the analysis is shown in the lower part of Table 10.7-26. It should be noted that $(AB)'$ cannot be distinguished from $(AC)'$ or $(BC)'$. This latter source of variation is sometimes called the uniqueness of the Greco-Latin square. In the partition of

the $CD \times \text{row}$ interaction, $(AB)'D$ cannot be distinguished from $(AC)'D$ or $(BC)'D$.

If all factors are fixed, all within-subject effects are tested by means of F ratios having $C \times \text{subjects}$ within group as a denominator. If interactions between the dimensions of the square are negligible, $(AB)'$ and $(AB)'D$ may be pooled with the experimental error. The F ratios for between-subject effects have subjects within groups as a denominator.

Table 10.7-27 Analysis of Variance for Briggs Data

Source	df	MS	F
Between subjects	79		
R Rows (sequences)	3	1,088	2.29
D Squares (initial learning)	4	8,975	18.85
DR	12	427	
Subjects w. groups	60	476	
Within subjects	240		
A Lists	3	1,668	10.80
B Overlearning	3	22,389	144.94
C Session	3	755	4.88
AD	12	122	
BD	12	346	2.24
CD	12	127	
$(AB)'$	3	152	
$(AB)'D$	12	114	
Error	180	154	

Illustrative Example. Briggs (1957) reports an experiment which is a special case of Plan 13. A 4×4 Greco-Latin square formed the basis for the plan. Groups of four subjects each were assigned to the rows of a Greco-Latin square. The letters of the square represented the lists (factor A) of syllables to be learned and the degree of overlearning (factor B). The columns of the square (factor C) represented the experimental session. There were five squares (factor D) in all—subjects in different squares received different amounts of initial learning. The same Greco-Latin square was used throughout; hence corresponding rows of the squares represent the same sequence of combinations of factors A and B .

A slightly modified analysis of variance for the Briggs data is given in Table 10.7-27.

Summary of Plans in Sec. 10.7

Plan 5

	a_1	a_2	a_3
G_1	b_3	b_1	b_2
G_2	b_1	b_2	b_3
G_3	b_2	b_3	b_1

Plan 6

	a_1	a_2	a_3
G_1	b_1	b_3	b_2
G_2	b_2	b_1	b_3
G_3	b_3	b_2	b_1

Plan 7

	a_1	a_2	a_3
G_1	bc_{11}	bc_{23}	bc_{32}
G_2	bc_{22}	bc_{31}	bc_{13}
G_3	bc_{33}	bc_{12}	bc_{21}

Plan 8

	a_1	a_2	a_3
c_1	G_1	b_1	b_2
	G_2	b_2	b_3
	G_3	b_3	b_1

	a_1	a_2	a_3
c_2	G_4	b_2	b_3
	G_5	b_1	b_2
	G_6	b_3	b_1

Plan 9

	a_1	a_2	a_3
c_1	G_1	b_2	b_3
	G_2	b_1	b_2
	G_3	b_3	b_1

	a_1	a_2	a_3
c_2	G_4	b_2	b_3
	G_5	b_1	b_2
	G_6	b_3	b_1

	a_1	a_2	a_3
c_3	G_7	b_2	b_3
	G_8	b_1	b_2
	G_9	b_3	b_1

Plan 10

	a_1	a_2	a_3
d_1	G_1	c_1	b_3
	G_2	c_2	b_2
	G_3	c_3	b_1

	a_1	a_2	a_3
d_2	G_4	c_1	b_2
	G_5	c_2	b_1
	G_6	c_3	b_3

Plan 11

	a_1	a_2	a_3
G_1	bc_{12}	bc_{31}	bc_{23}
G_2	bc_{33}	bc_{22}	bc_{11}
G_3	bc_{21}	bc_{13}	bc_{32}

	a_1	a_2	a_3
G_4	bc_{32}	bc_{21}	bc_{13}
G_5	bc_{23}	bc_{12}	bc_{31}
G_6	bc_{11}	bc_{33}	bc_{22}

	a_1	a_2	a_3
G_7	bc_{22}	bc_{11}	bc_{33}
G_8	bc_{13}	bc_{32}	bc_{21}
G_9	bc_{31}	bc_{23}	bc_{12}

Plan 12

	c_1	c_2
	b_1	b_2
	b_3	b_3
G_1	a_2	a_1
G_2	a_3	a_2
G_3	a_1	a_3

	c_1	c_2
	b_1	b_2
	b_3	b_3
a_2	a_1	a_3
a_3	a_2	a_1
a_1	a_3	a_2

Plan 13

	c_1	c_2	c_3
d_1	G_1	ab_{12}	ab_{31}
	G_2	ab_{33}	ab_{22}
	G_3	ab_{21}	ab_{13}

	c_1	c_2	c_3
d_2	G_4	ab_{12}	ab_{31}
	G_5	ab_{33}	ab_{22}
	G_6	ab_{21}	ab_{13}

CHAPTER 11

Analysis of Covariance

11.1 General Purpose

There are two general methods for controlling variability due to experimental error—direct and statistical. Direct control includes such methods as grouping the experimental units into homogeneous strata or blocks, increasing the uniformity of the conditions under which the experiment is run, and increasing the accuracy of the measurements. Replicated experiments, randomized block designs, repeated-measure designs, split-plot designs, incomplete-block designs—these designs use the direct-control principle to increase the precision of the experiment.

In this chapter, designs which use an indirect, or statistical, control (1) to increase the precision of the experiment and (2) to remove potential sources of bias in the experiment will be discussed. The latter objective is one that is particularly important in situations where the experimenter cannot assign the individual units at random to the experimental conditions. Statistical control is achieved by measuring one or more concomitant variates in addition to the variate of primary interest. The latter variate will be termed the criterion, or simply the variate; the concomitant variates will be called covariates. Measurements on the covariates are made for the purpose of adjusting the measurements on the variate.

For example, suppose that the purpose of an experiment is to determine the effect of various methods of extinction upon some kind of learned response. The variate in this experiment may be a measure of extinction; the covariate may be a measure associated with the degree of learning at the start of the extinction trials. As another example, suppose that the purpose of an experiment is to measure the effect of various stress situations upon blood pressure. In this case a measure of blood pressure under a condition of no stress may be the covariate. As still another example, suppose that the purpose of an experiment is to evaluate the effect of electrical stimulation on the weight of denervated muscle. The weight of the corresponding normal muscle may serve as a covariate in this type of experiment.

Direct control and statistical control may, of course, be used simultaneously within the same experiment. One or more variates may be under direct control; one or more variates may be under statistical control.

Suppose that the criterion means in a single-factor experiment are designated

$$\bar{Y}_1, \bar{Y}_2, \dots, \bar{Y}_k,$$

and that the covariate means are designated

$$\bar{X}_1, \bar{X}_2, \dots, \bar{X}_k.$$

Primary interest lies in differences among the \bar{Y}_j 's. Suppose that differences in the \bar{X}_j 's are due to sources of variation related to the \bar{Y}_j 's but not directly related to the treatment effects. If this is the case, then more precise information on the treatment effects may be obtained by adjusting the \bar{Y}_j 's for the effect of the \bar{X}_j 's. Suppose that the adjusted means on the criterion are designated

$$\bar{Y}'_1, \bar{Y}'_2, \dots, \bar{Y}'_k.$$

There are many ways of adjusting the criterion for the influence of the covariate. In some cases the adjustment may take the form of a simple difference between the corresponding means; in other cases the adjusted value may be the ratio of corresponding means. The appropriate adjustment for the influence of the covariate may, in some cases, be determined by experience gained from prior experimentation. More generally, the average effect of an increase of 1 unit in the covariate upon the variate is given by some form of regression analysis. Although the form of the regression need not be linear, only the linear case will be considered in this chapter. In terms of a linear regression, an adjusted criterion mean has the following form,

$$\bar{Y}'_j = \bar{Y}_j - b(\bar{X}_j - \bar{X}),$$

where b is an estimate of β , the population linear-regression coefficient. The estimate of β is obtained from the data of the experiment.

In terms of the basic linear model underlying the analysis of variance,

$$\bar{Y}_j = \mu + \beta(\bar{X}_j - \bar{X}) + \tau_j + \bar{\epsilon}_j,$$

$$\bar{Y}'_j = \mu + \tau_j + \bar{\epsilon}_j.$$

Thus differences between the adjusted criterion means are free of the linear effect of the covariate. If the effect due to the covariate is essentially linear, then this kind of adjustment process is adequate; otherwise the adequacy of the adjustment is questionable. The $\bar{\epsilon}_j$ term in the linear model includes the nonlinear variation in the covariate as well as all other sources of variation which are not under direct or statistical control. If the covariate measures the effects of some environmental source of variation that tends to inflate the experimental error, removal of the

influence of the covariate is equivalent to statistical control on the experimental error.

The change in experimental error due to this adjustment process depends upon the linear correlation between the variate and the covariate. If this correlation is equal to ρ , and if the experimental error per unit, disregarding the covariate, is σ_y^2 , then the experimental error after the adjustment is

$$\sigma_y^2(1 - \rho^2) \left[1 + \frac{1}{f_e - 2} \right],$$

where f_e represents the degrees of freedom for the estimation of σ_y^2 . If, instead of using a covariance adjustment, the experimental units could be grouped into blocks (equivalently, into strata or classes) which are homogeneous with respect to the covariate, and if the relation between X and Y is linear, then the effect of this kind of blocking is to reduce the experimental error from

$$\sigma_y^2 \quad \text{to} \quad \sigma_y^2(1 - \rho^2).$$

Thus, when regression is linear, covariance adjustment is approximately as effective as stratification with respect to the covariate. However, if the regression is not linear and a linear adjustment is used, then stratification will generally provide greater reduction in the experimental error. In a real sense, stratification is a function-free regression scheme.

When the covariate is actually affected by the treatment, the adjustment process removes more than an error component from the criterion; it also removes part of the treatment effect. An analysis of variance on the covariate will sometimes throw some light on whether or not the treatments are affecting the covariate. Apart from its function in reducing experimental error, the covariance adjustment may also be used to remove bias due to the covariate in studies where the experimenter must work with intact groups.

To illustrate this latter point, suppose that the criterion means for treatments j and m may be expressed in terms of the following linear model:

$$\bar{Y}_j = \mu_y + \tau_j + \beta(\mu_{x_j} - \mu_x) + \varepsilon_j,$$

$$\bar{Y}_m = \mu_y + \tau_m + \beta(\mu_{x_m} - \mu_x) + \varepsilon_m.$$

Then the difference between the criterion means has the expected value

$$E(\bar{Y}_j - \bar{Y}_m) = (\tau_j - \tau_m) + \beta(\mu_{x_j} - \mu_{x_m}).$$

The expected value of this difference is not a function of treatment effects alone—there is a bias which is a function of the difference between the covariate means and the magnitude of the regression coefficient. If the experimental groups are matched with respect to the covariate, or if $\beta = 0$, the bias disappears. However, if there are differences in the covariate

which cannot be subject to direct control and if $\beta \neq 0$, then the bias may be removed by a covariance adjustment. Direct control over this kind of bias is generally not possible when the experimenter is forced to work with intact groups. In such situations, statistical control of relevant sources of variation is the only method available to the experimenter.

In cases where differences between the covariate means \bar{X}_i and \bar{X}_j are relatively large, adjustment for this kind of bias is generally not based upon statistically sound principles. Such adjustments often require extrapolations which are either beyond the range of the data or in regions of the range where there are few observations. At best covariance adjustments for initial biases on the covariate are poor substitutes for direct controls. A more complete discussion of this latter point is given by Cochran (1957). An excellent summary of the wide variety of uses for the analysis of covariance is contained in a special issue of *Biometrics* (1957, 13, no. 3), which is devoted entirely to this topic.

11.2 Single-factor Experiments

The use of covariates to increase the precision of an experiment or to remove sources of bias is possible for all the designs that have been considered in earlier chapters. The principles underlying adjustment for covariates will be discussed in some detail for the case of a single-factor experiment. The generalization to other designs is relatively direct. The analysis of variance for designs which include covariates is frequently referred to as the analysis of covariance.

Consider a single-factor experiment in which there are k treatments. The notation that will be used for this case is summarized in Table 11.2-1. It will be assumed that there are n experimental units assigned to each of the treatment conditions. Two measurements are made on each of the experimental units—a measurement on the criterion, Y_{ij} , and a measurement on the covariate, X_{ij} . (Only the case of a single covariate will be considered in this section; the generalization to multiple covariates is direct.)

In the notation given in this table, T_{xx} represents n times the variation among the covariate means, E_{xx} represents the pooled within-class variation on the covariate, and S_{xx} represents the total variation for the covariate. T_{yy} , E_{yy} , and S_{yy} are, respectively, the analogous sources of variation for the criterion. The symbol T_{xy} represents n times the covariation between the covariate and criterion means, E_{xy} represents the pooled within-class covariation between the covariate and the criterion, and S_{xy} represents the total covariation between the covariate and the criterion.

Based upon the data for the n observations within treatment class j , the best estimate of the linear regression coefficient for the prediction of the criterion from the covariate is

$$b_j = \frac{E_{xyj}}{E_{xxj}};$$

Table 11.2-1 Notation for the Analysis of Covariance

Treatment 1		...	Treatment j		...	Treatment k	
Y_{11} Y_{21} . . . Y_{n1}	X_{11} X_{21} . . . X_{n1}		Y_{1j} Y_{2j} . . . Y_{nj}	X_{1j} X_{2j} . . . X_{nj}		Y_{1k} Y_{2k} . . . Y_{nk}	X_{1k} X_{2k} . . . X_{nk}
Sum Mean	T_{y_1} \bar{Y}_1	...	T_{y_j} \bar{Y}_j	T_{x_j} \bar{X}_j	...	T_{y_k} \bar{Y}_k	T_{x_k} \bar{X}_k
$T_{xx} = n\Sigma(\bar{X}_j - \bar{X})^2$ $E_{xxj} = \Sigma_i (X_{ij} - \bar{X}_j)^2$ $E_{xx} = \Sigma_j E_{xxj}$ $S_{xx} = T_{xx} + E_{xx} = \Sigma\Sigma(X_{ij} - \bar{X})^2$			$T_{xy} = n\Sigma(\bar{X}_j - \bar{X})(\bar{Y}_j - \bar{Y})$ $E_{xyj} = \Sigma_i (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j)$ $E_{xy} = \Sigma_j E_{xyj}$ $S_{xy} = T_{xy} + E_{xy} = \Sigma\Sigma(X_{ij} - \bar{X})(Y_{ij} - \bar{Y})$			$\Sigma T_{y_j} = G_y$ $\Sigma T_{x_j} = G_x$	
						$T_{yy} = n\Sigma(\bar{Y}_j - \bar{Y})^2$ $E_{yyj} = \Sigma_i (Y_{ij} - \bar{Y}_j)^2$ $E_{yy} = \Sigma_j E_{yyj}$ $S_{yy} = T_{yy} + E_{yy} = \Sigma\Sigma(Y_{ij} - \bar{Y})^2$	

b_j is called a within-class regression coefficient. The corresponding within-class product-moment correlation is

$$r_{xy_j} = \frac{b_j \sqrt{E_{xx_j}}}{\sqrt{E_{yy_j}}} = \frac{E_{xy_j}}{\sqrt{E_{xx_j}} \sqrt{E_{yy_j}}}.$$

If it can be assumed that

$$\beta_1 = \cdots = \beta_j = \cdots = \beta_k = \beta,$$

i.e., that the regression coefficients within each of the treatment classes are homogeneous, then within-class information from all the treatment classes may be pooled to provide a single estimate of the parameter β . This pooled estimate of β is

$$b = \frac{E_{xy}}{E_{xx}}.$$

Homogeneity of within-class regression is one of the fundamental assumptions underlying the analysis of covariance.

Adjusted criterion means have the general form

$$\bar{Y}'_j = \bar{Y}_j - b(\bar{X}_j - \bar{X}),$$

where b is the pooled within-class regression coefficient. The variation of \bar{Y}'_j is not, however, used in making an over-all test of the hypothesis that $\sigma^2_\tau = 0$. Because of sampling error in the estimation of β , and because the values of \bar{Y}'_j are not statistically independent (adjustments are correlated), a valid F test on the over-all hypothesis that $\sigma^2_\tau = 0$ requires a somewhat roundabout method for computing the numerator of the F ratio.

Suppose that the treatment classes are disregarded and a single linear regression equation is computed for the entire set of kn pairs of observations. Assume that this regression equation is

$$Y''_{ij} = b''(X_{ij} - \bar{X}) + \bar{Y}.$$

The residual sum of squares about this regression line may be shown to be

$$\Sigma \Sigma (Y_{ij} - Y''_{ij})^2 = S_{yy} - (S^2_{xy}/S_{xx}).$$

If there are n observations in each of the k treatment classes, there are kn statistically independent experimental units. The degrees of freedom for the variation of the residuals are $kn - 2$.

The regression equation based upon pooled within-class information has the form

$$Y'_j = b(X_{ij} - \bar{X}_j) + \bar{Y}_j.$$

The sum of squares of residuals about this regression line may be shown to be

$$\Sigma \Sigma (Y_{ij} - Y'_j)^2 = E_{yy} - \frac{E_{xy}^2}{E_{xx}}.$$

This source of variation has $k(n - 1) - 1$ degrees of freedom.

The reduced sum of squares due to treatments is, by definition, the difference between these two residuals, i.e.,

$$T_{yyR} = \left(S_{yy} - \frac{S_{xy}^2}{S_{xx}} \right) - \left(E_{yy} - \frac{E_{xy}^2}{E_{xx}} \right).$$

Since $S_{yy} - E_{yy} = T_{yy}$, the reduced sum of squares for treatments becomes

$$T_{yyR} = T_{yy} - \frac{S_{yy}^2}{S_{xx}} + \frac{E_{xy}^2}{E_{xx}}.$$

To summarize, the residual variation about the over-all regression line has been divided into two parts. One part is a function of within-class residual variation; the second part, the reduced sum of squares due to treatments, is a function of between-class variation.

Table 11.2-2 Analysis of Variance and Analysis of Covariance

	Source	SS	df	MS
(i)	Total Error	S_{yy} E_{yy}	$kn - 1$ $k(n - 1)$	MS_{error}
	Treatments	$T_{yy} = S_{yy} - E_{yy}$	$k - 1$	MS_{treat}
(ii)	Total Error	$S'_{yy} = S_{yy} - (S_{xy}^2/S_{xx})$ $E'_{yy} = E_{yy} - (E_{xy}^2/E_{xx})$	$kn - 2$ $k(n - 1) - 1$	MS'_{error}
	Treat (reduced)	$T_{yyR} = S'_{yy} - E'_{yy}$	$k - 1$	MS'_{treat}
<div style="display: flex; justify-content: space-around; align-items: center;"> <div style="text-align: center;"> <p>(iii) $F = \frac{MS_{\text{treat}}}{MS_{\text{error}}}$</p> <p>$[(k - 1), k(n - 1)]$</p> </div> <div style="text-align: center;"> <p>$F = \frac{MS'_{\text{treat}}}{MS'_{\text{error}}}$</p> <p>$[(k - 1), k(n - 1) - 1]$</p> </div> </div>				

This partition is given in part ii of Table 11.2-2. By way of contrast, the analysis of variance is shown in part i. It is seen that the analysis of variance partitions the total criterion variation into between- and within-class sources. In contrast, the analysis of covariance partitions the residuals from an over-all linear regression into between- and within-class sources.

The linear model in this analysis has the form

$$Y'_{ij} = Y_{ij} - \beta(X_{ij} - \bar{X}) = \mu + \tau_j + e_{ij}.$$

A test on the hypothesis that $\sigma_{\tau}^2 = 0$ makes use of the statistic

$$F = \frac{MS'_{\text{treat}}}{MS'_{\text{error}}}.$$

When the hypothesis being tested is true, numerator and denominator may be shown to be independent estimates of a common residual variance.

This F ratio has the general form of an F test in regression analysis. This form is

$$F = \frac{(D'' - D)/(k - 1)}{D/[k(n - 1) - 1]},$$

where D'' is the variation of the residuals about the over-all regression line and D is the variation of the residuals about the pooled within-class line.

It is of interest to indicate the relationship between variation due to adjusted treatment means, $\bar{Y}'_j = \bar{Y}_j - b(\bar{X}_j - \bar{X})$, and the reduced variation due to treatments. (The variation due to the \bar{Y}'_j 's will be denoted by $T_{yy.A}$.) Toward this end, a third regression equation needs to be introduced—the regression equation determined from the k pairs of means. This latter regression is called the between-class regression. It has the form

$$\bar{Y}'_j = b'''(\bar{X}_j - \bar{X}) + \bar{Y},$$

where b''' , the between-class regression coefficient, is

$$b''' = \frac{T_{xy}}{T_{xx}}.$$

This regression equation provides the best linear fit for predicting \bar{Y}_j from \bar{X}_j . The variation of the residuals of the \bar{Y}_j about this regression line is

$$n\Sigma(\bar{Y}_j - \bar{Y}'_j)^2 = T_{yy} - \frac{T_{xy}^2}{T_{xx}}.$$

Cochran (1957) has shown that

$$\begin{aligned} (1) \quad T_{yy.A} &= T_{yy} - 2bT_{xy} + b^2T_{xx} \\ &= T_{yy} - bb'''T_{xx} + b^2T_{xx}. \end{aligned}$$

From the relation (1) it may be shown that T_{yyR} and $T_{yy.A}$ are related as follows:

$$(2) \quad T_{yyR} = T_{yy.A} - \frac{T_{xx}^2(b''' - b)^2}{T_{xx} + E_{xx}}.$$

The relationship in (2) shows that T_{yyR} will always be less than $T_{yy.A}$. The latter is often used as an approximation for T_{yyR} . The relationship given in (1) provides a simple method for computing $T_{yy.A}$. If the treatments do not affect the covariate, $T_{yy.A}$ will generally be close to T_{yyR} .

A test on the difference between two adjusted treatment means has the form

$$F = \frac{(\bar{T}'_j - \bar{T}'_m)^2}{MS'_e \left[\frac{2}{n} + \frac{(\bar{X}_j - \bar{X}_m)^2}{E_{xx}} \right]}.$$

The degrees of freedom for this F ratio are 1 and $k(n - 1) - 1$, the latter being the degrees of freedom for MS'_e . Note that the variance of the difference between two adjusted means depends upon the magnitude of $(\bar{X}_j - \bar{X}_m)^2$. The average variance of the difference between two adjusted

means, averaged over all values of $(\bar{X}_j - \bar{X}_m)^2$, is

$$\frac{2MS'_e}{n} \left[1 + \frac{T_{xx}/(k-1)}{E_{xx}} \right].$$

This average variance of a difference between adjusted means may be used in making tests between several pairs or all possible pairs.

Ratios of various sums of squares are readily translated into correlation coefficients. The square of the within-class correlation is

$$r_{w, \text{class}}^2 = \frac{E_{xy}^2/E_{xx}}{E_{yy}}.$$

This correlation coefficient measures the goodness of fit of the within-class regression lines, using the regression coefficient b , to the individual observations. The square of the between-class correlation is

$$r_{\text{between class}}^2 = \frac{T_{xy}^2/T_{xx}}{T_{yy}}.$$

This correlation measures the goodness of fit of the between-class regression line to the criterion means.

Tests on the Assumptions Underlying the Analysis of Covariance. All the assumptions underlying the usual analysis-of-variance approach are also required in the analysis of covariance. In addition, there are assumptions about the regression effects. First, it is assumed that treatment effects and regression effects are additive. Implicit in this is the assumption that regressions are homogeneous. Second, it is assumed that the residuals are normally and independently distributed with zero means and the same variance. Implicit in this is the assumption that the proper form of regression equation has been fitted. If a linear regression is used when the true regression is curvilinear, then the assumptions made with respect to the residuals will generally not hold.

Evidence from the usual analysis of variance indicates that F tests in the analysis of covariance are robust with respect to the violation of the two assumptions, normality and homogeneity of the residual variance. The effect of nonhomogeneity of within-class regression, which is analogous to lack of additivity, has not been studied. There is an internal check that can be made on the assumption that the within-class regressions are homogeneous. Toward this end, the error (within-class) variation may be subdivided into the following parts:

Source	SS	df
Error	$E_{yy} - (E_{xy}^2/E_{xx})$	$k(n-1) - 1$
S_1	$E_{yy} - \Sigma(E_{xy_j}^2/E_{xx_j})$	$k(n-2)$
S_2	$\Sigma(E_{xy_j}^2/E_{xx_j}) - (E_{xy}^2/E_{xx})$	$k-1$

S_1 may be shown to be the variation of residuals of the form

$$\sum_i (Y_{ij} - Y_{ij}^*)^2,$$

where

$$Y_{ij}^* = b_j(X_{ij} - \bar{X}_j) + \bar{Y}_j.$$

This last regression equation is determined entirely from the data within treatment class j . (Each treatment class has its own regression line.) Within each class the variation due to the residuals has $n - 2$ degrees of freedom; the residuals from the k classes will have $k(n - 2)$ degrees of freedom.

S_2 may be shown to be a function of $\Sigma(b_j - b)^2$, the variation of the within-class regression coefficients about the pooled within-class regression coefficient. The larger this source of variation, the less reasonable it is to assume that the corresponding population regressions are equal. A test of this latter hypothesis, i.e., the hypothesis that $\beta_1 = \cdots = \beta_j = \cdots = \beta_k$, has the form

$$F = \frac{S_2/(k - 1)}{S_1/k(n - 2)}.$$

When the hypothesis being tested is true, this F ratio has a sampling distribution which can be approximated by an F distribution having $k - 1$ and $k(n - 2)$ degrees of freedom.

Source	SS	df
Treatments (reduced)	$T_{yy} - (S_{xy}^2/S_{xx}) + (E_{xy}^2/E_{xx})$	$k - 1$
S_3	$T_{yy} - (T_{xy}^2/T_{xx})$	$k - 2$
S_4	$(T_{xy}^2/T_{xx}) + (E_{xy}^2/E_{xx}) - (S_{xy}^2/S_{xx})$	1

The reduced sum of squares may also be partitioned into two parts. S_3 is a measure of the variation of the treatment means, \bar{Y}_j , about the between-class regression line. The larger this source of variation, the larger the deviation from linearity of the between-class regression. S_4 is a function of $(b'' - b)^2$, the squared deviation of the between-class regression coefficient and the pooled within-class regression coefficient.

If the within-class regression is linear, and if the covariate is not affected by the treatments, it is reasonable to expect that the between-class regression will also be linear. A test of the hypothesis that the between-class regression is linear has the form

$$F = \frac{S_3/(k - 2)}{MS'_{\text{error}}}.$$

When the hypothesis being tested is true, this F ratio has a sampling distribution which is approximated by an F distribution having $k - 2$ and

$k(n - 1) - 1$ degrees of freedom. If this regression does not prove to be linear, interpretation of the adjusted treatment means becomes difficult.

A test of the hypothesis that $\beta_{\text{between class}} = \beta_{\text{w. class}}$ has the form

$$F = \frac{S_4}{\text{MS}'_{\text{error}}}$$

The degrees of freedom for this statistic are 1 and $k(n - 1) - 1$. Note that $\beta_{\text{between class}}$ is estimated by b''' , whereas $\beta_{\text{w. class}}$ is estimated by b .

In the analysis of covariance, the residual variation about an over-all regression line which is fitted to the entire set of data, disregarding the treatment classes, is what is partitioned. An approximate test on the linearity of this over-all regression is given by

$$F = \frac{(S_2 + S_3 + S_4)/2(k - 1)}{S_1/k(n - 2)}$$

The degrees of freedom for this F ratio are $2(k - 1)$ and $k(n - 2)$.

The four different regression lines that have been used along the way are summarized in Table 11.2-3.

Table 11.2-3 Regression Equations in the Analysis of Covariance

	Regression equation	Residual variation
Within class j	$Y_{ij}^* = b_j(X_{ij} - \bar{X}_j) + \bar{Y}_j$	$E'_{yyj} = E_{yyj} - (E_{xyj}^2/E_{xxj})$
Pooled within class	$Y'_{ij} = b(X_{ij} - \bar{X}_j) + \bar{Y}_j$	$E'_{yy} = E_{yy} - (E_{xy}^2/E_{xx})$
Between class	$\bar{Y}_j'' = b''(\bar{X}_j - \bar{X}) + \bar{Y}$	$T'_{yy} = T_{yy} - (T_{xy}^2/T_{xx})$
Over-all	$Y_{ij}' = b''(X_{ij} - \bar{X}) + \bar{Y}$	$S'_{yy} = S_{yy} - (S_{xy}^2/S_{xx})$

11.3 Numerical Example of Single-factor Experiment

Computational symbols that will be used in the numerical example are summarized in Table 11.3-1. The symbols in parts i and ii are required only if tests on the assumptions underlying the analysis of covariance are made. Symbols required for the over-all analysis of covariance are summarized in parts iii and iv.

The numerical data in Table 11.3-2 will be used to illustrate the computational procedures. Suppose that the $k = 3$ treatments represent different methods of training. The experimenter is not at liberty to assign the subjects at random to the different methods of training; he is required to use groups that are already formed. However, from all available information, the groups were not originally selected on the basis of variables that are considered directly relevant to the study; for most practical purposes the groups can be considered as random samples from a common population.

There are $n = 7$ subjects in each of the groups. The groups are assigned at random to training methods. Before the subjects are trained under the method to which they are assigned, they are given a common aptitude test.

Table 11.3-1 Computational Formulas for Analysis of Covariance

(i)	$(2x_j) = \sum_i X_{ij}^2$	$(2xy_j) = \sum_i X_{ij} Y_{ij}$	$(2y_j) = \sum_i Y_{ij}^2$
	$(3x_j) = T_{xj}^2/n$	$(3xy_j) = T_{xj} T_{yj}/n$	$(3y_j) = T_{yj}^2/n$
(ii)	$E_{xx_1} = (2x_1) - (3x_1)$	$E_{xy_1} = (2xy_1) - (3xy_1)$	$E_{yy_1} = (2y_1) - (3y_1)$
	$E_{xx_k} = (2x_k) - (3x_k)$	$E_{xy_k} = (2xy_k) - (3xy_k)$	$E_{yy_k} = (2y_k) - (3y_k)$
(iii)	$(1x) = G_x^2/kn$	$(1xy) = G_x G_y/kn$	$(1y) = G_y^2/kn$
	$(2x) = \Sigma(2x_j)$	$(2xy) = \Sigma(2xy_j)$	$(2y) = \Sigma(2y_j)$
	$= \Sigma X^2$	$= \Sigma XY$	$= \Sigma Y^2$
	$(3x) = \Sigma(3x_j)$	$(3xy) = \Sigma(3xy_j)$	$(3y) = \Sigma(3y_j)$
	$= (\Sigma T_{xj}^2)/n$	$= (\Sigma T_{xj} T_{yj})/n$	$= (\Sigma T_{yj}^2)/n$
(iv)	$S_{xx} = (2x) - (1x)$	$S_{xy} = (2xy) - (1xy)$	$S_{yy} = (2y) - (1y)$
	$E_{xx} = (2x) - (3x)$	$E_{xy} = (2xy) - (3xy)$	$E_{yy} = (2y) - (3y)$
	$T_{xx} = (3x) - (1x)$	$T_{xy} = (3xy) - (1xy)$	$T_{yy} = (3y) - (1y)$
(v)		$S'_{yy} = S_{yy} - (S_{xy}^2/S_{xx})$	
		$E'_{yy} = E_{yy} - (E_{xy}^2/E_{xx})$	
		$T'_{yyR} = S'_{yy} - E'_{yy}$	

The scores on this test define the covariate measure X . After the training is completed, the subjects are given a common achievement test over the material covered in the training. The score on the latter test is the criterion measure Y . For example, the first subject in treatment class 1 made a score of $X = 3$ on the aptitude test and a score of $Y = 6$ on the achievement test.

In part i of this table, the entries under treatment 1 are as follows:

$$\begin{aligned} \sum_i X_{ij} &= 15 & \sum_i Y_{ij} &= 31 \\ \sum_i X_{ij}^2 &= 41 & \sum_i Y_{ij}^2 &= 147 \\ \sum_i X_{ij} Y_{ij} &= 75 \end{aligned}$$

In words, these entries are the sum, the sum of the squares, and the sum of the products of the pairs of observations under treatment 1. The totals to the right of part i are as follows:

$$\begin{aligned} \sum_j \left(\sum_i X_{ij} \right) &= 58 & \sum_j \left(\sum_i Y_{ij} \right) &= 131 \\ \sum_j \left(\sum_i X_{ij}^2 \right) &= 196 & \sum_j \left(\sum_i Y_{ij}^2 \right) &= 881 \\ \sum_j \left(\sum_i X_{ij} Y_{ij} \right) &= 398 \end{aligned}$$

Table 11.3-2 Numerical Example

		Treat 1		Treat 2		Treat 3		
		X	Y	X	Y	X	Y	
(i)		3	6	4	8	3	6	
		1	4	5	9	2	7	
		3	5	5	7	2	7	
		1	3	4	9	3	7	
		2	4	3	8	4	8	
		1	3	1	5	1	5	
		4	6	2	7	4	7	
	$\Sigma()$	15	31	24	53	19	47	Totals
	$\Sigma()^2$	41	147	96	413	59	321	58 131
	ΣXY	75		191		132		196 881
								398

		Total			
(ii)	E_{xx_j}	8.86	13.71	7.43	30.00 = E_{xx}
	E_{xy_j}	8.57	9.29	4.43	22.29 = E_{xy}
	E_{yy_j}	9.72	11.71	5.43	26.86 = E_{yy}

(iii)	$(1x) = (58)^2/21 = 160.19$				$(1y) = (131)^2/21 = 817.19$			
	$(2x) = 196$				$(2y) = 881$			
	$(3x) = (15^2 + 24^2 + 19^2)/7$				$(3y) = (31^2 + 53^2 + 47^2)/7$			
	$= 166.00$				$= 854.14$			
	$(1xy) = (58)(131)/21 = 361.81$							
	$(2xy) = 398$							
	$(3xy) = [(15)(31) + (24)(53) + (19)(47)]/7 = 375.71$							

(iv)	$S_{xx} = (2x) - (1x) = 35.81$				$S_{xy} = (2xy) - (1xy) = 36.19$			
	$E_{xx} = (2x) - (3x) = 30.00$				$E_{xy} = (2xy) - (3xy) = 22.29$			
	$T_{xx} = (3x) - (1x) = 5.81$				$T_{xy} = (3xy) - (1xy) = 13.90$			
	$S_{yy} = (2y) - (1y) = 63.81$							
	$E_{yy} = (2y) - (3y) = 26.86$							
	$T_{yy} = (3y) - (1y) = 36.95$							

These entries are the sums of the corresponding entries under the treatment headings.

The entries in part ii are within-class error terms for each of the classes. For example,

$$E_{xx_1} = 41 - (15^2/7) = 8.86,$$

$$E_{xy_1} = 75 - [(15)(31)/7] = 8.57,$$

$$E_{yy_1} = 147 - (31^2/7) = 9.72.$$

At the right of part ii are the totals for the entries in the corresponding rows.

Computational symbols used in the over-all analysis are given in part iii. These symbols are defined in part iii of Table 11.3-1.

To test the hypothesis of homogeneity of within-class regression, from part ii of Table 11.3-2 one computes

$$\begin{aligned}\sum \frac{E_{xyj}^2}{E_{xxj}} &= \frac{8.57^2}{8.86} + \frac{9.29^2}{13.71} + \frac{4.43^2}{7.43} \\ &= 8.29 + 6.29 + 2.64 = 17.22.\end{aligned}$$

The variation of the individual observations about the unpooled within-class regression lines is

$$S_1 = E_{yy} - \sum \frac{E_{xyj}^2}{E_{xxj}} = 26.86 - 17.22 = 9.64.$$

The variation of the individual within-class regression coefficients about the pooled within-class regression coefficient is

$$S_2 = \sum \frac{E_{xyj}^2}{E_{xxj}} - \frac{E_{xy}^2}{E_{xx}} = 17.22 - 16.56 = .66.$$

The statistic used in the test for homogeneity of within-class regression is

$$F = \frac{S_2/(k-1)}{S_1/k(n-2)} = \frac{(.66)/2}{(9.64)/15} = .52.$$

The critical value for a .10-level test is $F_{.90}(2,15) = 2.70$. The experimental data do not contradict the hypothesis of homogeneity of within-class

Table 11.3-3 Analysis of Variance

Source	SS	df	MS	F
Total	$S_{yy} = 63.81$	20		
Error	$E_{yy} = 26.86$	18	1.49	
Treatments	$T_{yy} = 36.95$	2	18.48	12.40

regression. The outcome of this test, in part, does not rule against pooling the within-class regressions. The appropriateness of the covariance model for the data justifies, in the long run, such pooling procedures.

An analysis of variance on the criterion variable alone is summarized in Table 11.3-3. Although the F test here indicates statistically significant differences in the treatment means, $F_{.99}(2,18) = 6.01$, direct interpretation of this test is difficult since the treatment classes appear to have different covariate means. A question might be raised about differences between treatments after adjustment is made for the linear trend in the relationship between the criterion and the covariate. Phrased in other words, do differences between the criterion means remain after a statistical adjustment has been made for the effects of the covariate? In a sense, the analysis of covariance attempts to approximate the situation in which each of the

treatment groups is equated on the covariate; this approximation procedure assumes that a linear model is appropriate.

The analysis of covariance is summarized in Table 11.3-4. The F ratio in this table provides a test of the hypothesis that $\sigma_\tau^2 = 0$, after the criterion data have been adjusted for the linear trend on the covariate. The critical

Table 11.3-4 Analysis of Covariance

Source	SS	df	MS	F
Total	$S'_{yy} = 27.24$	19		
Error	$E'_{yy} = 10.30$	17	.61	
Treatments	$T_{yyR} = 16.94$	2	8.47	13.89

value for a .01-level test in this case is $F_{.99}(2,17) = 6.11$. Thus the experimental data indicate statistically significant differences between the criterion scores for the groups even after adjustment is made for the linear effect of the covariate.

The adjusted means are given in Table 11.3-5. In making multiple tests between these means, the average effective error per unit is that given in this

Table 11.3-5 Adjusted Treatment Means

$b = E_{xy}/E_{xx} = .743$				
$\bar{X}_j \dots$	2.14	3.43	2.71	$\bar{X} = 2.76$
$\bar{X}_j - \bar{X} \dots$	-.62	.67	-.05	
$\bar{Y}_j \dots$	4.43	7.57	6.71	$\bar{Y} = 6.24$
$\bar{Y}'_j = \bar{Y}_j - b(\bar{X}_j - \bar{X}) \dots$	4.89	7.07	6.75	$\bar{Y}' = 6.24$

$$\begin{aligned}
 MS'_{\text{error(effective)}} &= MS'_{\text{error}} \left[1 + \frac{T_{xx}/(k-1)}{E_{xx}} \right] \\
 &= .61 \left[1 + \frac{5.81/2}{30.00} \right] = .67
 \end{aligned}$$

table. In making tests which fall in the a priori category, a somewhat different error term is used. For example, to test the hypothesis that $\tau_2 = \tau_3$,

$$F = \frac{(7.07 - 6.75)^2}{.61 \left[\frac{2}{7} + \frac{(3.43 - 2.71)^2}{30.00} \right]} = \frac{.1024}{.1848} = .55.$$

The critical value for a .01-level test in this case is $F_{.99}(1,17) = 8.40$. Had the average effective error been used, this ratio would be

$$F = \frac{(7.07 - 6.75)^2}{2(.67)/7} = \frac{.1024}{.1914} = .54.$$

The critical value used in the previous test would also be appropriate in this latter test.

It is of interest to examine the magnitude of the individual within-class correlations. The squares of these correlations are as follows:

$$r_1^2 = \frac{E_{xy_1}^2/E_{xx_1}}{E_{yy_1}} = .85, \quad r_2^2 = \frac{E_{xy_2}^2/E_{xx_2}}{E_{yy_2}} = .54, \quad r_3^2 = \frac{E_{xy_3}^2/E_{xx_3}}{E_{yy_3}} = .49.$$

The square of the pooled within-class correlation is

$$r_w^2 = \frac{E_{xy}^2/E_{xx}}{E_{yy}} = .62.$$

In terms of this latter correlation, the error term in the analysis of covariance is given by

$$\begin{aligned} MS'_{\text{error}} &= MS_{\text{error}}(1 - r_w^2) \left[1 + \frac{1}{k(n-1) - 2} \right] \\ &= 1.49(.383)(1 + \frac{1}{16}) = .61. \end{aligned}$$

This is the value for the error mean square in Table 11.3-4. Thus the reduction in the error term (increase in precision) is primarily a function of the magnitude of r_w^2 . The larger the latter correlation, the greater the reduction in the error term.

The reduced sum of squares for treatments, T_{yyR} , depends in part upon the magnitude of r_w^2 and in part upon the magnitude of the squared between-class correlation. The latter is

$$r_b^2 = \frac{T_{xy}^2/T_{xx}}{T_{yy}} = .90.$$

When r_b is large relative to r_w , the reduction in the treatment variation can be relatively larger than the reduction in the error variation. When this occurs, the F ratio in the analysis of covariance will actually be smaller than the corresponding F ratio in the analysis of variance. The latter finding would generally indicate that bias due to the covariate had inflated the F ratio in the original analysis of variance.

When r_b is negative and r_w is positive, T_{yyA} will always be larger than T_{yy} . The reduced sum of squares, T_{yyR} , may in this case also be larger

than T_{yy} . When this happens, the F ratio in the original analysis of variance is deflated because of the bias due to the covariate.

If the treatment classes are disregarded, the square of the over-all correlation between the covariate and the criterion is

$$r_{\text{total}}^2 = \frac{S_{xy}^2 / S_{xx}}{S_{yy}} = .57.$$

Rather than using a covariance analysis for the data in Table 11.3-2, the experimenter might have attempted to use the covariate as a classification or stratification factor. If this were done, the experiment would be analyzed as a two-factor experiment. The data would have the form given in Table 11.3-6. In this particular case the cell frequencies would be quite small;

Table 11.3-6 Use of Covariate as a Classification Factor

Covariate classification	Treatment		
	1	2	3
5		9,7	
4	6	8,9	8,7
3	6,5	8	6,7
2	4	7	7,7
1	4,3,3	5	5

further, there would be no entries in some of the cells. If each of the resulting cell frequencies is relatively large, say 5 or more, this type of stratification on the covariate is generally to be preferred to the analysis of covariance. Often neighboring ordered categories having small frequencies may be grouped in order to build up the cell frequencies.

Unequal Cell Frequencies. If the number of subjects in each treatment class is not constant, the only changes in the computational formulas are as follows:

$$\begin{aligned} (3x_j) &= \frac{T_{xj}^2}{n_j} & (3xy_j) &= \frac{T_{xj} T_{yj}}{n_j} & (3y_j) &= \frac{T_{yj}^2}{n_j} \\ (3x) &= \Sigma(3x_j) & (3xy) &= \Sigma(3xy_j) & (3y) &= \Sigma(3y_j) \end{aligned}$$

The degrees of freedom for this case are as follows:

Source	df
S'_{yy}	$(\Sigma n_j) - 2$
E'_{yy}	$[\Sigma(n_j - 1)] - 1$
T'_{yy}	$k - 1$

With these changes, the computational procedures outlined in Table 11.3-1 may be used.

11.4 Factorial Experiment

The analysis of covariance for a factorial experiment is a direct generalization of the corresponding analysis for a single-factor experiment. Assuming a $p \times q$ factorial experiment having n observations in each cell, the model is as follows:

$$Y'_{ij} = Y_{ij} - \beta(X_{ij} - \bar{X}) = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \varepsilon_{ij}.$$

That is, an observation, adjusted for the effect of the covariate, estimates the parameters in the usual analysis of variance. If the covariate is ignored, its effect augments variation due to experimental error; the latter includes all uncontrolled sources of variation. It will be assumed throughout this section that A and B represent fixed factors.

The observations on the covariate and the criterion within cell ab_{jk} of the experiment are represented as follows:

X	Y
X_{1jk}	Y_{1jk}
X_{2jk}	Y_{2jk}
\dots	\dots
X_{njk}	Y_{njk}

In each of the pq cells in the experiment there are n pairs of observations. The following notation denotes the various sums needed in the analysis:

AB_{xjk} = sum of covariate measures in cell ab_{jk} .

AB_{yjk} = sum of criterion measures in cell ab_{jk} .

A_{xj} = sum of all covariate measures at level a_j .

A_{yj} = sum of all criterion measures at level a_j .

B_{xk} = sum of all covariate measures at level b_k .

B_{yk} = sum of all criterion measures at level b_k .

G_x = sum of all covariate measures.

G_y = sum of all criterion measures.

The means corresponding to the sets of sums defined above are obtained by dividing the respective sums by the number of experimental units over which the sum is taken. For example,

$$\overline{AB}_{xjk} = \frac{AB_{xjk}}{n}, \quad \bar{A}_{yj} = \frac{A_{yj}}{nq}, \quad \bar{G}_y = \frac{G_y}{npq}.$$

The variation of the covariate within cell ab_{jk} is

$$E_{xxjk} = \sum_i (X_{ijk} - \overline{AB}_{xjk})^2.$$

The corresponding variation of the criterion is

$$E_{yyjk} = \sum_i (Y_{ijk} - \overline{AB}_{yjk})^2.$$

The covariation within this cell is

$$E_{xy_{jk}} = \sum_i (X_{ijk} - \overline{AB}_{x_{jk}})(Y_{ijk} - \overline{AB}_{y_{jk}}).$$

From these data, the within-cell regression coefficient is

$$b_{jk} = \frac{E_{xy_{jk}}}{E_{xx_{jk}}}.$$

There will be pq such regression coefficients.

If the within-cell variations of both the covariate and criterion are homogeneous, and if the within-cell covariations are homogeneous, then corresponding sources of variation may be pooled. These pooled estimates of the within-cell sources of variation are

$$E_{xx} = \Sigma\Sigma(E_{xx_{jk}}), \quad E_{yy} = \Sigma\Sigma(E_{yy_{jk}}), \quad E_{xy} = \Sigma\Sigma(E_{xy_{jk}}).$$

The estimate of the within-cell regression coefficient based upon the pooled data is

$$b = \frac{E_{xy}}{E_{xx}}.$$

The pooled within-class regression line has the following form:

$$Y'_{ijk} = b(X_{ijk} - \overline{AB}_{x_{jk}}) + \overline{AB}_{y_{jk}}.$$

The variation of the observations about this set of regression lines is

$$E'_{yy} = \Sigma\Sigma\Sigma(Y_{ijk} - Y'_{ijk})^2 = E_{yy} - \frac{E_{xy}^2}{E_{xx}}.$$

This source of variation provides an estimate of the variation due to experimental error in the analysis of covariance; its form is identical to the corresponding term in a single-factor experiment.

The variation due to the main effect of treatment A for the covariate is

$$A_{xx} = nq\Sigma(\bar{A}_{x_j} - \bar{G}_x)^2.$$

The corresponding variation for the criterion is

$$A_{yy} = nq\Sigma(\bar{A}_{y_j} - \bar{G}_y)^2.$$

The covariation associated with the main effect of factor A is

$$A_{xy} = nq\Sigma(\bar{A}_{x_j} - \bar{G}_x)(\bar{A}_{y_j} - \bar{G}_y).$$

Just as in the case of a single-factor experiment, the adjusted variation due to main effects of factor A is obtained in a rather roundabout manner. One first combines the following sources of variation:

$$\frac{E_{xx}}{A_{xx} + E_{xx}} \quad \frac{E_{xy}}{A_{xy} + E_{xy}} \quad \frac{E_{yy}}{A_{yy} + E_{yy}}$$

The residual variation corresponding to this last line is

$$(A + E)'_{yy} = (A_{yy} + E_{yy}) - \frac{(A_{xy} + E_{xy})^2}{A_{xx} + E_{xx}}.$$

The adjusted variation for the main effect of factor A is

$$A'_{yy} = (A + E)'_{yy} - E'_{yy} = A_{yy} - \frac{(A_{xy} + E_{xy})^2}{A_{xx} + E_{xx}} + \frac{E_{xy}^2}{E_{xx}}.$$

The variation due to the main effect of factor B is given by a procedure analogous to that used in obtaining the main effect of factor A . Thus,

$$\begin{aligned} B_{xx} &= np\Sigma(\bar{B}_{xk} - \bar{G}_x)^2, \\ B_{yy} &= np\Sigma(\bar{B}_{yk} - \bar{G}_y)^2, \\ B_{xy} &= np\Sigma(\bar{B}_{xk} - \bar{G}_x)(\bar{B}_{yk} - \bar{G}_y). \end{aligned}$$

Similarly,

$$\frac{E_{xx}}{B_{xx} + E_{xx}} \quad \frac{E_{xy}}{B_{xy} + E_{xy}} \quad \frac{E_{yy}}{B_{yy} + E_{yy}}$$

The adjusted variation due to the main effect of factor B is

$$B'_{yy} = (B + E)'_{yy} - E'_{yy} = B_{yy} - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} + \frac{E_{xy}^2}{E_{xx}}.$$

The sources of variation associated with the AB interaction are defined as follows:

$$\begin{aligned} AB_{xx} &= n\Sigma\Sigma(\bar{AB}_{xjk} - \bar{A}_{xj} - \bar{B}_{xk} + \bar{G}_x)^2, \\ AB_{yy} &= n\Sigma\Sigma(\bar{AB}_{yjk} - \bar{A}_{yj} - \bar{B}_{yk} + \bar{G}_y)^2, \\ AB_{xy} &= n\Sigma\Sigma(\bar{AB}_{xjk} - \bar{A}_{xj} - \bar{B}_{xk} + \bar{G}_x)(\bar{AB}_{yjk} - \bar{A}_{yj} - \bar{B}_{yk} + \bar{G}_y). \end{aligned}$$

To obtain the variation due to the interaction adjusted for the covariate, one first combines the following sources of variation:

$$\frac{E_{xx}}{AB_{xx} + E_{xx}} \quad \frac{E_{xy}}{AB_{xy} + E_{xy}} \quad \frac{E_{yy}}{AB_{yy} + E_{yy}}$$

The adjusted variation for this last line is

$$(AB + E)'_{yy} = (AB_{yy} + E_{yy}) - \frac{(AB_{xy} + E_{xy})^2}{AB_{xx} + E_{xx}}.$$

The variation due to AB , adjusted for the covariate, is

$$\begin{aligned} AB'_{yy} &= (AB + E)'_{yy} - E'_{yy} \\ &= AB_{yy} - \frac{(AB_{xy} + E_{xy})^2}{AB_{xx} + E_{xx}} + \frac{E_{xy}^2}{E_{xx}}. \end{aligned}$$

The analysis of variance is contrasted with the analysis of covariance in Table 11.4-1. Assuming all factors fixed, MS'_{error} is the proper denominator for F tests on main effects and the interaction. An approximation to the

Table 11.4-1 Summary of Analysis of Variance and Analysis of Covariance

Source	Analysis of variance			Analysis of covariance		
	SS	df	MS	SS	df	MS
A	A_{yy}	$p - 1$	MS_a	A'_{yy}	$p - 1$	MS'_a
B	B_{yy}	$q - 1$	MS_b	B'_{yy}	$q - 1$	MS'_b
AB	AB_{yy}	$(p - 1)(q - 1)$	MS_{ab}	AB'_{yy}	$(p - 1)(q - 1)$	MS'_{ab}
Error	E_{yy}	$pq(n - 1)$	MS_{error}	E'_{yy}	$pq(n - 1) - 1$	MS'_{error}

adjusted sums of squares for the main effects and interactions uses an expression analogous to (1) in Sec. 11.2. Thus,

$$A''_{yy} = A_{yy} - 2bA_{xy} + b^2A_{xx},$$

$$B''_{yy} = B_{yy} - 2bB_{xy} + b^2B_{xx},$$

$$AB''_{yy} = AB_{yy} - 2bAB_{xy} + b^2AB_{xx},$$

where $b = E_{xy}/E_{yy}$.

In cases where the treatments do not affect the covariate or in cases where bias due to the covariate is not large, the approximations given above will be quite close to the estimates obtained by the somewhat more laborious procedures described earlier.

The adjusted cell means have the following form:

$$\overline{AB'}_{ijk} = \overline{AB}_{xjk} - b(\overline{AB}_{xjk} - \overline{AB}_x).$$

The average effective error for the difference between two adjusted cell means is

$$\frac{2MS'_{\text{error}}}{n} \left[1 + \frac{AB_{xx}/(p - 1)(q - 1)}{E_{xx}} \right].$$

In making multiple comparisons between the cell means, the effective error per experimental unit is

$$MS'_{\text{error}} \left[1 + \frac{AB_{xx}/(p - 1)(q - 1)}{E_{xx}} \right].$$

The adjusted mean for level a_j is

$$\bar{A}'_{yj} = \bar{A}_{yj} + b(\bar{A}_{xj} - \bar{C}_x).$$

The variance of a difference between two adjusted means is

$$\frac{2MS'_{\text{error}}}{nq} \left[1 + \frac{(\bar{A}_{xj} - \bar{A}_{xm})^2}{E_{xx}} \right].$$

In making multiple tests among the \bar{A}_j 's, the average effective error per experimental unit may be taken as

$$MS'_{\text{error}} \left[1 + \frac{A_{xx}/(p-1)}{E_{xx}} \right].$$

An adjusted mean for level b_k is

$$\bar{B}'_{y_k} = \bar{B}_{y_k} + b(\bar{B}_{x_k} - \bar{G}_x).$$

The corresponding average experimental error per unit is

$$MS'_{\text{error}} \left[1 + \frac{B_{xx}/(q-1)}{E_{xx}} \right].$$

In making a test for homogeneity of within-cell regression, the error term may be partitioned as follows:

Source	df
$E_{yy} = E_{yy} - (E_{xy}^2/E_{xx})$	$pq(n-1) - 1$
$S_1 = E_{yy} - \Sigma\Sigma(E_{xy_{jk}}^2/E_{xx_{jk}})$	$pq(n-2)$
$S_2 = \Sigma\Sigma(E_{xy_{jk}}^2/E_{xx_{jk}}) - (E_{xy}^2/E_{xx})$	$pq - 1$

S_1 represents the variation of the observations about the individual within-cell regression lines. S_2 represents the variation of the individual cell regression coefficients about the pooled within-class regression coefficient. The test for homogeneity of within-class regression uses the statistic

$$F = \frac{S_2/(pq-1)}{S_1/pq(n-2)}.$$

It should be noted that the pooled within-class regression coefficient is used in obtaining \bar{A}'_y , \bar{B}'_j , and \bar{AB}'_y . Implicit in this adjustment process is the assumption that b is the appropriate regression coefficient for all these adjustments. In most experimental situations in the behavioral sciences this assumption very probably oversimplifies what is a more complex model. Caution in the use of covariates in factorial experiments is advised. The model for the analysis outlined in this section is a highly restrictive one. The experimenter should be aware that such designs exist, but he also should be aware of both their strengths and their weaknesses.

11.5 Computational Procedures for Factorial Experiment

Computational procedures for the analysis of covariance for a $p \times q$ factorial experiment having n observations per cell will be considered first. The case of unequal cell frequencies is considered later in this section, where

the equivalent of an unweighted-means analysis is outlined. Procedures in this latter case require relatively small changes from those to be given for the case of equal cell frequencies.

Assuming n pairs of observations in each of the pq cells, computational formulas for the sums of squares needed in the analysis of covariance are given in Table 11.5-1. In each case the range of summation is over all possible values of the total that is squared and summed. Symbols $(1x)$

Table 11.5-1 Computational Formulas for the Analysis of Covariance in a Factorial Experiment

(i)	$(1x) = G_x^2/npq$	$(1xy) = G_x G_y / npq$	$(1y) = G_y^2 / npq$
	$(2x) = \Sigma X^2$	$(2xy) = \Sigma XY$	$(2y) = \Sigma Y^2$
	$(3x) = (\Sigma A_x^2) / nq$	$(3xy) = (\Sigma A_x A_y) / nq$	$(3y) = (\Sigma A_y^2) / nq$
	$(4x) = (\Sigma B_x^2) / np$	$(4xy) = (\Sigma B_x B_y) / np$	$(4y) = (\Sigma B_y^2) / np$
	$(5x) = (\Sigma AB_x^2) / n$	$(5xy) = (\Sigma AB_x A_y) / n$	$(5y) = (\Sigma AB_y^2) / n$
(ii)	$A_{xx} = (3x) - (1x)$	$A_{xy} = (3xy) - (1xy)$	$A_{yy} = (3y) - (1y)$
	$B_{xx} = (4x) - (1x)$	$B_{xy} = (4xy) - (1xy)$	$B_{yy} = (4y) - (1y)$
	$AB_{xx} = (5x) - (3x)$	$AB_{xy} = (5xy) - (3xy)$	$AB_{yy} = (5y) - (3y)$
	$- (4x) + (1x)$	$- (4xy) + (1xy)$	$- (4y) + (1y)$
	$E_{xx} = (2x) - (5x)$	$E_{xy} = (2xy) - (5xy)$	$E_{yy} = (2y) - (5y)$
$E'_{yy} = E_{yy} - (E_{xy}^2 / E_{xx})$			
(iii)	$(A + E)_{yy}' = (A_{yy} + E_{yy}) - \frac{(A_{xy} + E_{xy})^2}{A_{xx} + E_{xx}}$		$A'_{yy} = (A + E)_{yy}' - E'_{yy}$
	$(B + E)_{yy}' = (B_{yy} + E_{yy}) - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}}$		$B'_{yy} = (B + E)_{yy}' - E'_{yy}$
	$(AB + E)_{yy}' = (AB_{yy} + E_{yy}) - \frac{(AB_{xy} + E_{xy})^2}{(AB_{xx} + E_{xx})}$		$AB'_{yy} = (AB + E)_{yy}' - E'_{yy}$

through $(5x)$ are those used in an analysis of variance on the covariate. Symbols $(1y)$ through $(5y)$ are those used in the usual analysis of variance on the criterion. Symbols $(1xy)$ through $(5xy)$ are used to estimate the covariances needed in the adjustment process.

A 2×3 factorial experiment having $n = 5$ observations per cell will be used to illustrate the computational procedures. Suppose that the $p = 2$ levels of factor A represent methods of instructing in teaching map reading, and suppose that the $q = 3$ levels of factor B represent three instructors. For purposes of the present analysis, both factors will be considered to be fixed. The covariate measure in this experiment is the score on an achievement test on map reading prior to the training; the criterion measure is the score on a comparable form of the achievement test after training is completed. Assume that intact groups of $n = 5$ subjects each are assigned at random to the cells of the experiment. Suppose that data obtained from this experiment are those given in the upper part of Table 11.5-2.

Table 11.5-2 Numerical Example

Method	b_1 Instr. 1		b_2 Instr. 2		b_3 Instr. 3	
	X	Y	X	Y	X	Y
a_1	40	95	30	85	50	90
	35	80	40	100	40	85
	40	95	45	85	40	90
	50	105	40	90	30	80
	45	100	40	90	40	85
a_2	50	100	50	100	45	95
	30	95	30	90	30	85
	35	95	40	95	25	75
	45	110	45	90	50	105
	30	88	40	95	35	85

AB summary:

	b_1		b_2		b_3		Total	
	X	Y	X	Y	X	Y	X	Y
a_1	210	475	195	450	200	430	605	1355
a_2	190	488	205	470	185	445	580	1403
	400	963	400	920	385	875	1185	2758

$(1x) = 46,808$	$(1xy) = 108,941$	$(1y) = 253,552$
$(2x) = 48,325$	$(2xy) = 110,065$	$(2y) = 255,444$
$(3x) = 46,828$	$(3xy) = 108,901$	$(3y) = 253,629$
$(4x) = 46,822$	$(4xy) = 109,008$	$(4y) = 253,939$
$(5x) = 46,895$	$(5xy) = 108,979$	$(5y) = 254,019$

In this table, for example, the first subject under method a_1 and instructor b_2 has scores of 30 and 85, respectively, on the covariate and the criterion. In symbols,

$$X_{121} = 30, \quad Y_{121} = 85.$$

An AB summary table appears under the observed data. There are two entries in each cell of this table—one represents the sum of the observations on the covariate, the other the sum for the criterion. For example, the sum of the $n = 5$ observations on the covariate under treatment combination ab_{12} is

$$AB_{x_{12}} = 30 + 40 + 45 + 40 + 40 = 195.$$

The corresponding sum for the criterion data is

$$AB_{y_{12}} = 85 + 100 + 85 + 90 + 90 = 450.$$

The entries in the total columns at the right of the AB summary table are the sums of corresponding entries in the rows. For example,

$$A_{x_1} = \sum_k AB_{x_1k} = 210 + 195 + 200 = 605.$$

The corresponding sum for the criterion data is

$$A_{y_1} = \sum_k AB_{y_1k} = 475 + 450 + 430 = 1355.$$

The total of the first column in the summary table is

$$B_{x_1} = \sum_j AB_{x_{j1}} = 210 + 190 = 400.$$

The corresponding sum for the criterion is

$$B_{y_1} = \sum_j AB_{y_{j1}} = 475 + 488 = 963.$$

The grand totals for the covariate and the criterion are

$$G_x = \sum A_{x_j} = \sum B_{x_k} = 1185,$$

$$G_y = \sum A_{y_j} + \sum B_{y_k} = 2758.$$

The computational symbols in the lower part of Table 11.5-2 are defined in part i of Table 11.5-1. The only symbols requiring special comment are those in the center column. These entries are obtained as follows:

$$(1xy) = \frac{(1185)(2758)}{30},$$

$$(2xy) = (40)(95) + (35)(80) + \cdots + (50)(105) + (35)(85),$$

$$(3xy) = \frac{(605)(1355) + (580)(1403)}{15},$$

$$(4xy) = \frac{(400)(963) + (400)(920) + (385)(875)}{10},$$

$$(5xy) = \frac{(210)(475) + (190)(488) + \cdots + (185)(445)}{5}.$$

The basic data for all of these symbols except $(2xy)$ are obtained from the AB summary table.

Sums of squares and sums of products are given in Table 11.5-3. Computational formulas for these terms are given in parts ii and iii of Table 11.5-1. Note that it is possible for the entries that are used to obtain covariances to be either positive or negative. In this case the between-class covariation of the totals corresponding to the main effects of factor A is negative (-40). Inspection of the total columns at the right of the AB summary in Table 11.5-2 indicates that the higher criterion total is paired with the lower covariate total; hence the negative covariation.

Table 11.5-3 Summary Data for Numerical Example

$A_{xx} = 21$	$A_{xy} = -40$	$A_{yy} = 77$
$B_{xx} = 15$	$B_{xy} = 67$	$B_{yy} = 387$
$AB_{xx} = 52$	$AB_{xy} = 11$	$AB_{yy} = 3$
$E_{xx} = 1430$	$E_{xy} = 1086$	$E_{yy} = 1425$
1518	1124	1892
$E'_{yy} = 1425 - (1086^2/1430)$		
$= 600$		
$(A + E)'_{yy} = 748$	$A'_{yy} = 748 - 600 = 148$	
$(B + E)'_{yy} = 892$	$B'_{yy} = 892 - 600 = 292$	
$(AB + E)'_{yy} = 616$	$AB'_{yy} = 616 - 600 = 16$	

The analysis of variance for the criterion data is summarized in Table 11.5-4. This analysis disregards the presence of the covariate. Differences between the methods of training are tested by means of the statistic

$$F = \frac{77}{59.4} = 1.30.$$

(The instructor factor is considered to be fixed.) This test indicates no statistically significant difference between the methods in so far as the mean of the groups is concerned.

Table 11.5-4 Analysis of Variance

Source	SS	df	MS	F
A Methods	$A_{yy} = 77$	1	77	1.30
B Instructors	$B_{yy} = 387$	2	193.5	3.26
AB	$AB_{yy} = 3$	2	1.5	
Error	$E_{yy} = 1425$	24	59.4	
Total	1892	29		

$$F_{.95}(2,24) = 3.40$$

The analysis of covariance is summarized in Table 11.5-5. Note that the error mean square in this case is 26.1, compared with 59.4 in the case of the analysis of variance. Further note that the adjusted method mean square is 148, compared with 77 in the analysis of variance. This increase in the adjusted method variance is a function of the negative covariance for the between-method totals. A .05-level test on the methods in the analysis of covariance indicates statistically significant differences between the criterion means. Thus, when a linear adjustment is made for the effect of variation due to differences in prior experience in map reading, as measured by the covariate, there are statistically significant differences between the training methods.

Table 11.5-5 Analysis of Covariance

Source	SS	df	MS	F
A Methods	$A'_{yy} = 148$	1	148.0	5.67
B Instructors	$B'_{yy} = 292$	2	146.0	5.59
AB	$AB'_{yy} = 16$	2	8.0	
Error	$E'_{yy} = 600$	23	26.1	
		28		

$$F_{.95}(1,23) = 4.28; F_{.95}(2,23) = 3.44$$

An estimate of the square of the within-cell correlation is

$$r^2_{\text{within}} = \frac{E^2_{xy}/E_{xx}}{E_{yy}} = .58.$$

The mean square due to experimental error in the analysis of covariance is approximately

$$MS'_{\text{Error}} = (1 - r^2_{\text{within}})MS_{\text{Error}}.$$

The adjusted criterion means for factor *A* are given in Table 11.5-6. Note that the difference between the adjusted means is larger than the

Table 11.5-6 Adjusted Means

$b = E_{xy}/E_{xx} = .76$			
	Method 1	Method 2	Mean
\bar{A}_{x_j}	40.3	38.7	$39.5 = \bar{G}_x$
$\bar{A}_{x_j} - \bar{G}_x$.8	-.8	
\bar{A}_y	90.3	93.5	$91.9 = \bar{G}_y$
$\bar{A}'_y = \bar{A}_y - .76(\bar{A}_{x_j} - \bar{G}_x)$	89.7	94.1	

corresponding difference between the unadjusted means. Had the covariance in this case been positive rather than negative, the difference between the adjusted means would have been smaller rather than larger than the difference between the unadjusted means.

It is of interest to compare the adjusted mean squares given in Table 11.5-5 with those that would be obtained by the approximation method described in the last section. The latter mean squares are as follows:

	SS	MS
$A''_y = A_{yy} - 2bA_{xy} + b^2A_{xx}$ $= 77 - 2(.76)(-40) + (.76^2)(21)$	150	150
$B''_y = B_{yy} - 2bB_{xy} - b^2B_{xx}$	309	154
$AB''_y = AB_{yy} - 2bAB_{xy} - b^2AB_{xx}$	16	8

These mean squares are slightly larger than the corresponding mean squares in Table 11.5-5.

Unequal Cell Frequencies. Under conditions in which an unweighted-means analysis is appropriate for an analysis of variance, there is an equivalent unweighted-means analysis appropriate for the analysis of covariance. With the definitions of the computational symbols given in Table 11.5-7, the procedures outlined in Table 11.5-1 may be used to obtain an unweighted-means analysis of covariance.

Table 11.5-7 Computational Formulas (Unequal Cell Frequencies)

$(1x) = G_x^2/N$	$(1xy) = G_x G_y/N$	$(1y) = G_y^2/N$
$(2x) = \Sigma X^2$	$(2xy) = \Sigma XY$	$(2y) = \Sigma Y^2$
$(3x) = \tilde{n}q(\Sigma \bar{A}_x^2)$	$(3xy) = \tilde{n}q(\Sigma \bar{A}_x \bar{A}_y)$	$(3y) = \tilde{n}q(\Sigma \bar{A}_y^2)$
$(4x) = \tilde{n}p(\Sigma \bar{B}_x^2)$	$(4xy) = \tilde{n}p(\Sigma \bar{B}_x \bar{B}_y)$	$(4y) = \tilde{n}p(\Sigma \bar{B}_y^2)$
$(5x) = \tilde{n}(\Sigma \bar{A}_x \bar{B}_x^2)$	$(5xy) = \tilde{n}(\Sigma \bar{A}_x \bar{B}_x \bar{A}_y)$	$(5y) = \tilde{n}(\Sigma \bar{A}_y \bar{B}_y^2)$

In this analysis it is assumed that the number of observations in each of the pq cells is approximately n , the harmonic mean of the cell frequencies. The latter is

$$\tilde{n} = \frac{pq}{\Sigma \Sigma (1/n_{jk})}$$

The total number of observations in the experiment is designated N . To avoid excessive rounding error, several decimal places should be carried on each of the means during the intermediate computational stages. The computational symbols defined in Table 11.5-7 are algebraically equivalent to those in part i of Table 11.5-1 when there are equal cell frequencies.

Illustrative Applications. Learning and conditioning experiments have made relatively extensive use of the analysis of covariance. Prokasy, Grant, and Myers (1958) report a typical application of the analysis of covariance in this context. The purpose of their experiment was to study the effect of stimulus intensity (four levels) and intertrial interval (three levels) upon acquisition and extinction of eyelid conditioning. The criterion was the proportion of conditioned responses in units of the arcsine transformation. In the acquisition phase of the study, the score on the first day's trials served as the covariate for analyzing the results of the second day's trials. In the extinction phase, the combined scores during the first and second days' trials during acquisition served as the covariate. The analysis of covariance during the acquisition phase was used to clarify interpretation of the trend of the effects. It was found that the intertrial interval had an effect during day 1, and the effect continued during day 2; however, the latter effect was almost entirely predictable from that observed on day 1. Zimbardo and Miller (1958) reported an experiment in which covariance analysis is employed for essentially the same purpose.

Payne (1958) reports an experiment that is a special case of a $2 \times 2 \times 2$ factorial experiment. The plan of this experiment may be represented as follows:

Method of initial learning	Method of relearning	Drugs	
		c_1	c_2
a_1	b_1	G_{111}	G_{112}
	b_2	G_{121}	G_{122}
a_2	b_1	G_{211}	G_{212}
	b_2	G_{221}	G_{222}

The covariate in this study was the logarithm of the number of trials required for initial learning; the criterion was the logarithm of the number of trials required for relearning. For example, the subjects in group G_{121} learned initially under method a_1 . Relearning was done under method b_2 while the subjects were under the influence of drug c_1 . The use of covariance analysis in this case is to adjust the relearning data for the linear effect of the initial learning.

Cotton, Lewis, and Metzger (1958) report an experiment which is a $3 \times 2 \times 2$ factorial. The type of apparatus in which a behavior pattern was acquired defined one factor. A second factor was defined by the type of apparatus in which the behavior pattern was extinguished. The time of restriction in a goal box defined the third factor. The covariate for the extinction data was the score on the last five trials of the acquisition phase.

11.6 Factorial Experiment—Repeated Measures

The design to be considered in this section is the analogue of the split-plot design in agricultural research. The design may be represented schematically as shown at the left of Table 11.6-1. Assume that there are n subjects in each of the groups. At the right is the usual analysis of variance for this design when there is no covariate.

Table 11.6-1 Factorial Experiment, Repeated Measures

					Source	SS	df
	b_1	b_2	\cdots	b_q	Between subjects		$np - 1$
a_1	G_1	G_1	\cdots	G_1	A	A_{yy}	$p - 1$
a_2	G_2	G_2	\cdots	G_2	Subj w. gp	P_{yy}	$p(n - 1)$
\vdots					Within subjects		$np(q - 1)$
\vdots					B	B_{yy}	$q - 1$
\vdots					AB	AB_{yy}	$(p - 1)(q - 1)$
a_p	G_p	G_p	\cdots	G_p	Residual	E_{yy}	$p(q - 1)(n - 1)$
							$npq - 1$

Adjustment procedures depend upon whether or not the between- and within-subject regression coefficients can be considered homogeneous. In the first part of the discussion that follows these regression coefficients are not assumed homogeneous.

Aside from the matter of homogeneity of the between- and within-subject regressions, two cases of this design need to be distinguished. In case (1) there is a single covariate measure associated with all the criterion scores for an individual. The data for subject i may be represented as follows:

		b_1	b_2	\cdots	b_q
Subject i	X_i	Y_{i1}	Y_{i2}	\cdots	Y_{iq}

Here the covariate score X_i is a measure taken before the administration of any of the treatments. Hence the same X_i is paired with all criterion scores on subject i . In contrast, for case (2) the covariate measure is taken just before, just after, or simultaneously with the criterion measure. The data for subject i in this case may be represented as follows:

	b_1		b_2		\cdots	b_q	
Subject i	X_{i1}	Y_{i1}	X_{i2}	Y_{i2}	\cdots	X_{iq}	Y_{iq}

Thus for case (2) each criterion measure on subject i is paired with a unique covariate measure. Case (1) may be considered as a special case of case (2) in which all the X_{ij} 's for subject i are equal. Hence computational procedures for case (2) may be used for case (1).

Under case (2), both the between-subject (whole-plot) comparisons and the within-subject (split-plot) comparisons are adjusted for the effect of the covariate. Under case (1), only the between-subject (whole-plot) comparisons are adjusted for the effect of the covariate—the within-subject (split-plot) comparisons will all have adjustments which are numerically equal to zero.

The notation that will be used is essentially that defined in Sec. 11.4. Two additional symbols are required.

$P_{x_{i(j)}}$ = sum of the q observations on the covariate for subject i in group G_j .

$P_{y_{i(j)}}$ = sum of the q observations on the criterion for subject i in group G_j .

The following variations and covariations are associated with differences between subjects within the groups:

$$\begin{aligned} P_{xx} &= q \sum \sum (\bar{P}_{x_{i(j)}} - \bar{A}_{x_j})^2, \\ P_{xy} &= q \sum \sum (\bar{P}_{x_{i(j)}} - \bar{A}_{x_j})(\bar{P}_{y_{i(j)}} - \bar{A}_{y_j}), \\ P_{yy} &= q \sum \sum (\bar{P}_{y_{i(j)}} - \bar{A}_{y_j})^2. \end{aligned}$$

The linear regression equation for the subjects within-group data has the following form:

$$\bar{P}'_{y_{i(j)}} = b_p(\bar{P}_{x_{i(j)}} - \bar{A}_{x_j}) + \bar{A}_{y_j},$$

Table 11.6-2 Covariance Analysis for Design in Table 11.6-1

Source	X^2	XY	Y^2	Adjusted variation	df
<i>A</i> Subj w. gp	A_{xx}	A_{xy}	A_{yy}	$A'_{yy} = (A_{yy} + P_{yy}) - \frac{(A_{xy} + P_{xy})^2}{A_{xx} + P_{xx}} - P'_{yy}$ $P'_{yy} = P_{yy} - (P_{xy}^2 / P_{xx})$	$p - 1$ $p(n - 1) - 1$
	P_{xx}	P_{xy}	P_{yy}		
<i>B</i> <i>AB</i>	B_{xx}	B_{xy}	B_{yy}	$B'_{yy} = (B_{yy} + E_{yy}) - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} - E'_{yy}$ $AB'_{yy} = (AB_{yy} + E_{yy}) - \frac{(AB_{xy} + E_{xy})^2}{AB_{xx} + E_{xx}} - E'_{yy}$ $E'_{yy} = E_{yy} - (E_{xy}^2 / E_{xx})$	$q - 1$ $(p - 1)(q - 1)$ $p(q - 1)(n - 1) - 1$
	AB_{xx}	AB_{xy}	AB_{yy}		
Residual	E_{xx}	E_{xy}	E_{yy}		

Note: If a single covariate measure is associated with all scores on a given subject, the following sums of squares will be equal to zero:

$$B_{xx}, AB_{xx}, E_{xx}, B_{xy}, AB_{xy}, E_{xy}.$$

where b_p is given by

$$b_p = \frac{P_{xy}}{P_{xx}}.$$

The variation of the residuals about this regression line is

$$P'_{yy} = P_{yy} - \frac{P_{xy}^2}{P_{xx}}.$$

The mean square corresponding to this latter variation is the adjusted error for between-subject effects.

The variations and covariations associated with the treatment effects are estimated by the same procedures as those used in Sec. 11.4. The error terms for within-subject effects in this design have the following form:

$$\text{Residual } (x) = \text{within cell } (x) - P_{xx}.$$

Thus,

$$\begin{aligned} E_{xx} &= \Sigma(X_{ijk} - \overline{AB}_{xjk})^2 - P_{xx}, \\ E_{xy} &= \Sigma[X_{ijk}Y_{ijk} - (\overline{AB}_{xjk})(\overline{AB}_{yjk})] - P_{xy}, \\ E_{yy} &= \Sigma(Y_{ijk} - \overline{AB}_{yjk})^2 - P_{yy}. \end{aligned}$$

The regression coefficient for the within-subject effects is given by

$$b_w = \frac{E_{xy}}{E_{xx}}.$$

The residuals about the regression line for the within-subject effects have variation equal to

$$E'_{yy} = E_{yy} - \frac{E_{xy}^2}{E_{xx}}.$$

An outline of the analysis of covariance is given in Table 11.6-2.

Computational procedures for this design differ only slightly from those given in Table 11.5-1. One additional set of computational symbols is required for the case of repeated measures.

$$(6x) = \frac{(\Sigma P_x^2)}{q}, \quad (6xy) = \frac{\Sigma(P_x P_y)}{q}, \quad (6y) = \frac{(\Sigma P_y^2)}{q}.$$

In each case the summation is over all possible values of the totals. A more complete notation for these totals is $P_{x(ij)}$, each total being over q observations. In terms of the symbols defined in Table 11.5-1 and the set of symbols defined here,

$$\begin{aligned} P_{xx} &= (6x) - (3x), & P_{xy} &= (6xy) - (3xy), \\ E_{xx} &= (2x) - (5x) - (6x) + (1x), & E_{xy} &= (2xy) - (5xy) - (6xy) + (1xy), \\ P_{yy} &= (6y) - (3y), \\ E_{xy} &= (2y) - (5y) - (6y) + (1y). \end{aligned}$$

The variation associated with the treatment effects is defined in Table 11.5-1. The adjusted mean for level a_j has the form

$$\bar{A}'_{y_j} = \bar{A}_{y_j} - b_p(\bar{A}_{x_j} - \bar{G}_x).$$

The effective error variance for a difference between two adjusted means in this case is

$$s_p^2 \left[\frac{2}{nq} + \frac{(\bar{A}_{x_j} - \bar{A}_{x_m})^2}{P_{xx}} \right],$$

where s_p^2 is the adjusted between-subject mean square for error. An adjusted mean for level b_j has the form

$$\bar{B}'_{y_k} = \bar{B}_{y_k} - b_w(\bar{B}_{x_k} - \bar{G}_x).$$

The effective error variance for the difference between two adjusted means in this case is

$$s_w^2 \left[\frac{2}{np} + \frac{(\bar{B}_{x_k} - \bar{B}_{x_m})^2}{E_{xx}} \right],$$

where s_w^2 is the adjusted within-subject mean square for error.

An adjusted cell mean is given by

$$\bar{AB}'_{y_{jk}} = \bar{AB}_{y_{jk}} - b_p(\bar{A}_{x_j} - \bar{G}_x) - b_w(\bar{AB}_{x_{jk}} - \bar{A}_{x_j}).$$

The effective error for the difference between two adjusted cell means which are at the same level of factor A is estimated by

$$s_w^2 \left[\frac{2}{n} + \frac{(\bar{AB}_{x_{jk}} - \bar{AB}_{x_{jm}})^2}{E_{xx}} \right].$$

The difference between two adjusted cell means which are not at the same level of factor A has the following form:

$$\begin{aligned} \bar{AB}'_{y_{jk}} - \bar{AB}'_{y_{ms}} &= \bar{AB}_{y_{jk}} - \bar{AB}_{y_{ms}} - b_p(\bar{A}_{x_j} - \bar{A}_{x_m}) \\ &\quad - b_w(\bar{AB}_{x_{jk}} - \bar{A}_{x_j} - \bar{AB}_{x_{ms}} + \bar{A}_{x_m}). \end{aligned}$$

Since this difference involves both between- and within-subject effects, the effective error variance is somewhat more complex. The latter variance is estimated by

$$\frac{2[s_p^2 + (q-1)s_w^2]}{nq} + \frac{(\bar{A}_{x_j} - \bar{A}_{x_m})^2 s_p^2}{P_{xx}} + \frac{(\bar{AB}_{x_{jk}} - \bar{A}_{x_j} - \bar{AB}_{x_{ms}} + \bar{A}_{x_m})^2 s_w^2}{E_{xx}}.$$

If the regression coefficients for between-subject (β_p) and within-subject (β_w) effects are equal, the within-subject regression may be used throughout in making the adjustments. When the treatments do not affect the covariate, it is reasonable to expect that the between- and within-subject regressions will be equal. A test on the hypothesis that $\beta_p = \beta_w$ is given by

$$t' = \frac{b_p - b_w}{\sqrt{s_1^2 + s_2^2}},$$

where $s_1^2 = s_p^2/P_{xx}$ and $s_2^2 = s_w^2/E_{xx}$. s_1^2 and s_2^2 are the respective error variances for b_p and b_w . Since the variances in the denominator of this t' statistic will not in general be homogeneous, the sampling distribution of t' is not that of the usual t statistic. If the degrees of freedom for s_p^2 and s_w^2 are both larger than 20, the normal distribution $N(0,1)$ may be used to approximate the sampling distribution of t' . In other cases the sampling distribution of t' may be approximated by the usual t distribution with degrees of freedom f ,

$$f = \frac{(s_1^2 + s_2^2)^2}{(s_1^4/f_p) + (s_2^4/f_w)},$$

where f_p and f_w are the respective degrees of freedom for s_p^2 and s_w^2 .

When it can be assumed that $\beta_p = \beta_w$, the analysis of covariance has the form given in Table 11.6-3. All the adjustments for within-subject effects are identical to those given in Table 11.6-2. The adjustment procedures for the between-subject effects are, however, different. For purposes of making over-all tests, there is some indication that the adjustments for between-subject effects given at the bottom of this table are to be preferred to those indicated at the top. The two adjustment procedures are not algebraically equivalent.

Adjusted means for within-subject effects are identical to those given earlier in this section. The adjusted mean for level a_j now has the form

$$\bar{A}'_{y_j} = \bar{A}_{y_j} - b_w(\bar{A}_{x_j} - \bar{G}_x).$$

The error variance for the difference between \bar{A}'_{y_j} and \bar{A}'_{y_m} in this case is approximately

$$s_p'^2 \left[\frac{2}{nq} + \frac{(\bar{A}_{x_j} - \bar{A}_{x_m})^2}{E_{xx}} \right].$$

An adjusted cell mean in this case is

$$\bar{AB}'_{y_{jk}} = \bar{AB}_{y_{jk}} - b_w(\bar{AB}_{x_{jk}} - \bar{G}_x).$$

The error variance for the difference between two adjusted cell means which are at the same level of factor A is identical to that given earlier in this section—this difference is a within-subject effect. The error variance for the difference between two adjusted cell means which are not at the same level of A is approximately

$$\frac{2[s_p'^2 + (q-1)s_w^2]}{nq} + \frac{(\bar{AB}_{x_{jk}} - \bar{AB}_{x_{ms}})^2 s_w^2}{E_{xx}}.$$

Numerical Example. The data in Table 11.6-4 will be used to illustrate the computational procedures for this design. Disregarding the covariate, part i represents data that would be obtained in a 2×2 factorial experiment having repeated measures on factor B . There are four subjects under each level of factor A . Suppose that the covariate measure on a given subject is obtained before the administration of any of the treatments; then the

Table 11.6-3 Covariance Analysis When $\beta_p = \beta_w$

Source	Adjusted variation	df	MS
<i>A</i>	$A''_{yy} = (A_{yy} + E_{yy}) - \frac{(A_{xy} + E_{xy})^2}{A_{xx} + E_{xx}} - E'_{yy}$	$p - 1$	
Subj w. gp	$P''_{yy} = (P_{yy} + E_{yy}) - \frac{(P_{xy} + E_{xy})^2}{P_{xx} + E_{xx}} - E'_{yy}$	$p(n - 1)$	$S_p'^2$
<i>B</i>	$B'_{yy} = (B_{yy} + E_{yy}) - \frac{(B_{xy} + E_{xy})^2}{(B_{xx} + E_{xx})} - E'_{yy}$	$q - 1$	
<i>AB</i>	$AB'_{yy} = (AB_{yy} + E_{yy}) - \frac{(AB_{xy} + E_{xy})^2}{(AB_{xx} + E_{xx})} - E'_{yy}$	$(p - 1)(q - 1)$	
Residual	$E'_{yy} = E_{yy} - (E_{xy}^2 / E_{xx})$	$p(q - 1)(n - 1) - 1$	$S_w'^2$
<i>A</i>	$A''_{yy} = A_{yy} - 2b_w A_{xy} + b_w^2 A_{xx}$	$p - 1$	
Subj w. <i>A</i>	$P''_{yy} = P_{yy} - 2b_w P_{xy} + b_w^2 P_{xx}$	$p(n - 1)$	$S_p'^2$

covariate measure for a given subject will be a constant for both levels of factor B . Thus the data in Table 11.6-4 represent case (1) of this design.

Table 11.6-4 Numerical Example

	Person	b_1		b_2		Total	
		X	Y	X	Y	X	Y
(i)	a_1	3	10	3	8	6	18
		5	15	5	12	10	27
		8	20	8	14	16	34
		2	12	2	6	4	18
	a_2	1	15	1	10	2	25
		8	25	8	20	16	45
		10	20	10	15	20	35
		2	15	2	10	4	25
		39	132	39	95	78	227
		b_1		b_2		Total	
		X	Y	X	Y	X	Y
(ii)	a_1	18	57	18	40	36	97
	a_2	21	75	21	55	42	130
		39	132	39	95	78	227
(iii)	$(1x) = 380.25$	$(1xy) = 1106.62$		$(1y) = 3220.56$			
	$(2x) = 542$	$(2xy) = 1282$		$(2y) = 3609$			
	$(3x) = 382.50$	$(3xy) = 1119.00$		$(3y) = 3288.62$			
	$(4x) = 380.25$	$(4xy) = 1106.62$		$(4y) = 3306.12$			
	$(5x) = 382.50$	$(5xy) = 1119.00$		$(5y) = 3374.75$			
	$(6x) = 542.00$	$(6xy) = 1282.00$		$(6y) = 3516.50$			

An AB summary table for both the variate and covariate appears in part ii. With the exception of symbols $(6x)$, $(6xy)$, and $(6y)$, the computational symbols given in part iii are defined in part i of Table 11.5-1. Symbols containing the number (6) were defined earlier in this section. Not all the symbols in part iii are required in the analysis of data for case (1), since within-subject (split-plot) adjustments will be zero; to show, however, that such adjustments will be zero, the complete analysis will be illustrated.

The computation of the variation for the between- and within-subject effects is illustrated in Table 11.6-5. By inserting x , xy , or y in the symbols given at the left of part i, one obtains the variations given under the headings X^2 , XY , and Y^2 , respectively. For example,

$$\begin{aligned}
 A_{xx} &= (3x) - (1x) = 2.25, \\
 A_{xy} &= (3xy) - (1xy) = 12.38, \\
 A_{yy} &= (3y) - (1y) = 68.06.
 \end{aligned}$$

The adjusted value for the between-subject error term is given by

$$P'_{yy} = P_{yy} - \frac{P_{xy}^2}{P_{xx}} = 227.88 - \frac{163.00^2}{159.50} = 61.30.$$

The adjusted value for the variation due to the main effect of factor A is given by

$$\begin{aligned} A'_{yy} &= (A_{yy} + P_{yy}) - \frac{(A_{xy} + P_{xy})^2}{A_{xx} + P_{xx}} - P'_{yy} \\ &= (68.06 + 227.88) - \frac{(12.38 + 163.00)^2}{2.25 + 159.50} - 61.30 \\ &= 44.48. \end{aligned}$$

Table 11.6-5 Analysis of Variance and Covariance

				X^2	XY	Y^2	$(Y')^2$
(i)	A	(3) - (1)		2.25	12.38	68.06	44.48
	Subj w. A	(6) - (3)		159.50	163.00	227.88	61.30
	B	(4) - (1)		0	0	85.56	
	AB	(5) - (3) - (4) + (1)		0	0	.57	
	Residual	(2) - (5) - (6) + (3)		0	0	6.37	

Source of variation		SS	df	MS	F
(ii)	A	68.06	1	68.06	1.79
	Subj w. A	227.88	6	37.98	
	B	85.56	1	85.56	80.72
	AB	.57	1	.57	
	Residual	6.37	6	1.06	
A (adj)		44.48	1	44.48	3.63
Subj w. A (adj)		61.30	5	12.26	

The analysis of variance as well as the analysis of covariance are summarized in part ii of Table 11.6-5. Only the between-subject (whole-plot) comparisons will have nonzero adjustments. If one were to compute the adjusted value for the main effect due to factor B , the adjusted value would be

$$\begin{aligned} B'_{yy} &= (B_{yy} + E_{yy}) - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} - E'_{yy} \\ &= (85.56 + 6.37) - \frac{(0 + 0)^2}{0 + 0} - 6.37 = 85.56 = B_{yy}. \end{aligned}$$

That these adjustments should be zero follows from the fact that the covariate measure is constant for all criterion measures on the same subject.

It should be noted that the error mean square for the between-subject (whole-plot) comparisons is 37.98 in the analysis of variance. In the analysis of covariance the corresponding error mean square is 12.26. Hence the covariate adjustment reduces the error mean square for between-subject comparisons from 37.98 to 12.26. This relatively large reduction in the error mean square is a function of the magnitude of the within-class correlation. The square of this latter correlation is estimated by

$$r_w^2 = \frac{P_{xy}^2 / P_{xx}}{P_{yy}} = .73.$$

The mean square due to the main effect of factor *A* is reduced from 68.06 to 44.48 in the adjustment process.

Under the analysis of variance, the *F* ratio for the test on the main effect of factor *A* is 1.79, with degrees of freedom 1 and 6; the corresponding *F* ratio under the analysis of covariance is 3.63, with degrees of freedom 1 and 5.

Another Numerical Example. The data in Table 11.6-6 will be used to illustrate a design in which there are distinct covariate measures for each observation. The design in this table may be considered as a special case of

Table 11.6-6 Numerical Example

	Subject	<i>b</i> ₁		<i>b</i> ₂		Total	
		<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>	<i>X</i>	<i>Y</i>
<i>a</i> ₁	1	3	8	4	14	7	22
	2	5	11	9	18	14	29
	3	11	16	14	22	25	38
		19	35	27	54	46	89
<i>a</i> ₂	4	2	6	1	8	3	14
	5	8	12	9	14	17	26
	6	10	9	9	10	19	19
		20	27	19	32	39	59
<i>a</i> ₃	7	7	10	4	10	11	20
	8	8	14	10	18	18	32
	9	9	15	12	22	21	37
		24	39	26	50	50	89
Total		63	101	72	136	135	237

(1 <i>x</i>) = 1012.50	(1 <i>xy</i>) = 1777.50	(1 <i>y</i>) = 3120.50
(2 <i>x</i>) = 1233	(2 <i>xy</i>) = 2004	(2 <i>y</i>) = 3495
(3 <i>x</i>) = 1022.83	(3 <i>xy</i>) = 1807.50	(3 <i>y</i>) = 3220.50
(4 <i>x</i>) = 1017.00	(4 <i>xy</i>) = 1795.00	(4 <i>y</i>) = 3188.56
(5 <i>x</i>) = 1034.33	(5 <i>xy</i>) = 1835.67	(5 <i>y</i>) = 3305.00
(6 <i>x</i>) = 1207.50	(6 <i>xy</i>) = 1964.00	(6 <i>y</i>) = 3397.50

a 3×2 factorial experiment with repeated measures on factor B . There are $n = 3$ subjects in each group.

Computational symbols obtained in part ii are defined in part i of Table 11.5-1. Symbol (6) has the following definition:

$$(6x) = \frac{(\sum P_x^2)}{q} = \frac{7^2 + 14^2 + \cdots + 18^2 + 21^2}{2} = 1207.50,$$

$$(6xy) = \frac{(\sum P_x P_y)}{q} = \frac{7(22) + 14(19) + \cdots + 21(37)}{2} = 1964.00,$$

$$(6y) = \frac{(\sum P_y^2)}{q} = \frac{22^2 + 29^2 + \cdots + 32^2 + 37^2}{2} = 3397.50.$$

The sums of squares and sums of products for the variations and covariations are summarized in part i of Table 11.6-7. The analysis of variance is summarized in part ii. The analysis of covariance is summarized in

Table 11.6-7 Analysis of Variance and Covariance

Table 11.3-7 Analysis of Variance and Covariance

			X^2	XY	Y^2
(i)	A	(3) - (1)	10.33	30.00	100.00
	Subj w. A	(6) - (3)	184.67	156.50	177.00
	B	(4) - (1)	4.50	17.50	68.06
	AB	(5) - (3) - (4) + (1)	7.00	10.67	16.44
	$B \times (\text{subj w. } A)$	(2) - (5) - (6) + (3)	14.00	11.83	13.00
	Total	(2) - (1)	220.50	226.50	374.50

	Source	SS	df	MS	F
(ii)	Between subjects				
	A	100.00	2	50.00	1.69
	Subj w. A	177.00	6	29.50	
	Within subjects				
	B	68.06	1	68.06	47.26
	AB	16.44	2	8.22	5.71
	$B \times (\text{subj w. } A)$	13.00	9	1.44	

	Source (adjusted)	SS	df	MS	F
(iii)	Between subjects				
	A'	54.26	2	27.13	3.06
	Subj w. A (P'_{yy})	44.37	5	8.87	
	Within subjects				
	B'	31.56	1	31.56	84.16
	AB'	2.33	2	1.16	3.09
	$B \times (\text{subj w. } A)$ (E'_{yy})	3.00	8	0.375	

part iii. The adjusted variation is defined in Table 11.6-2. For example,

$$E'_{yy} = E_{yy} - \frac{E_{xy}^2}{E_{xx}} = 13.00 - \frac{(11.83)^2}{14.00} = 3.00,$$

$$B'_{yy} = (B_{yy} + E_{yy}) - \frac{(B_{xy} + E_{xy})^2}{B_{xx} + E_{xx}} - E'_{yy}$$

$$= (68.06 + 13.00) - \frac{(17.50 + 11.83)^2}{4.50 + 14.00} - 3.00 = 31.56.$$

The between-subject effects in part iii are also adjusted in accordance with the procedures given in Table 11.6-2. The estimates of b_p and b_w are

$$b_p = \frac{P_{xy}}{P_{xx}} = \frac{156.50}{184.67} = .847,$$

$$b_w = \frac{E_{xy}}{E_{xx}} = \frac{11.83}{14.00} = .845.$$

In this case, inspection indicates that b_w may be used throughout in the adjustment process. If the adjustment of the between-subject effects is made in accordance with procedures given at the bottom of Table 11.6-3, the adjusted sums of squares are as follows:

$$A'_{yy} = A_{yy} - 2b_w A_{xy} + b_w^2 A_{xx}$$

$$= 100.00 - 2(.845)(30.00) + (.845)^2(10.33)$$

$$= 56.68,$$

$$P'_{yy} = P_{yy} - 2b_w P_{xy} + b_w^2 P_{xx}$$

$$= 177.00 - 2(.845)(156.50) + (.845)^2(184.67)$$

$$= 44.36.$$

The over-all F test for the hypothesis $\sigma_\alpha^2 = 0$ has the form

$$F = \frac{A'_{yy}/(p-1)}{P'_{yy}/p(n-1)} = \frac{56.68/2}{44.36/6} = 3.83.$$

It is noted that the error term for the within-subject effects is 1.44 in the analysis of variance and 0.375 in the analysis of covariance. For the between-subject effects the corresponding error terms are 29.50 and $44.36/6 = 7.39$.

Illustrative Application. Suppose that an experimenter is interested in evaluating the effect of two drugs upon blood pressure. Suppose further that the experimenter desires to evaluate effect upon blood pressure when a linear adjustment is made for the effect of the drugs upon pulse rate. The covariate, pulse rate, is in part affected by the drugs. In the design, a random sample of $2n$ subjects is divided at random into two groups, each of

size n . Subjects in group 1 receive drug 1 first. After the drug has had a chance to act, blood pressure and pulse rate are taken on each subject. After the effect of drug 1 has dissipated, subjects in group 1 are given drug 2. Blood pressure and pulse rate are then obtained under the latter drug. Subjects in group 2 are given the drugs in the reverse order. The design may be represented schematically as follows:

	Drug	
	b_1	b_2
a_1	G_1	G_1
a_2	G_2	G_2

The analysis of covariance for this design will permit the experimenter to evaluate the effect of the drugs upon blood pressure after a linear adjustment has been made for the effect of the drugs upon pulse rate. In this type of experiment, one or more levels of the drug factor may represent control conditions.

11.7 Multiple Covariates

The principles developed for the case of a single covariate generalize to more than one covariate. In this section this generalization will be considered in the setting of a single-factor experiment. However, multiple covariates may be used with any of the experimental designs for which there is a corresponding analysis of variance.

Suppose that there are two covariates X and Z , and a single criterion Y . The observations under treatment j in a single-factor experiment may be represented as follows:

Treatment j		
X_{1j}	Z_{1j}	Y_{1j}
X_{2j}	Z_{2j}	Y_{2j}
\vdots	\vdots	\vdots
\vdots	\vdots	\vdots
X_{ij}	Z_{ij}	Y_{ij}
\vdots	\vdots	\vdots
\vdots	\vdots	\vdots
X_{nj}	Z_{nj}	Y_{nj}

In the analysis of covariance, the criterion data are adjusted for the linear effects of both X and Z through use of a multiple regression equation. The multiple regression equation obtained from the data in class j has the form

$$Y_{ij}'' = b_{y.x_j}(X_{ij} - \bar{X}_j) + b_{y.z_j}(Z_{ij} - \bar{Z}_j) + \bar{Y}_j.$$

If the within-class regressions are homogeneous, pooled within-class regression coefficients may be computed. The multiple regression equation in terms of the pooled within-class regression coefficients has the form

$$Y'_{ij} = b_{y \cdot x}(X_{ij} - \bar{X}_j) + b_{y \cdot z}(Z_{ij} - \bar{Z}_j) + \bar{Y}_j.$$

The pooled within-class variances and covariances are defined as follows:

$$\begin{aligned} E_{xx} &= \Sigma \Sigma (X_{ij} - \bar{X}_j)^2, & E_{zz} &= \Sigma \Sigma (Z_{ij} - \bar{Z}_j)^2, \\ E_{xy} &= \Sigma \Sigma (X_{ij} - \bar{X}_j)(Y_{ij} - \bar{Y}_j), & E_{zy} &= \Sigma \Sigma (Z_{ij} - \bar{Z}_j)(Y_{ij} - \bar{Y}_j), \\ E_{yy} &= \Sigma \Sigma (Y_{ij} - \bar{Y}_j)^2, & E_{xz} &= \Sigma \Sigma (X_{ij} - \bar{X}_j)(Z_{ij} - \bar{Z}_j). \end{aligned}$$

In terms of these variances and covariances, the pooled within-class regression coefficients are given by

$$\begin{aligned} b_{y \cdot x} &= \frac{E_{zz}E_{xy} - E_{xz}E_{zy}}{d}, \\ b_{y \cdot z} &= \frac{-E_{xz}E_{xy} + E_{xx}E_{zy}}{d}, \end{aligned}$$

where $d = E_{xx}E_{zz} - E_{xz}^2$. The sum of the squares of the residuals about this regression line is

$$E'_{yy} = E_{yy} - b_{y \cdot x}E_{xy} - b_{y \cdot z}E_{zy}.$$

The treatment variations and covariations are defined as follows:

$$\begin{aligned} T_{xx} &= n \Sigma (\bar{T}_x - \bar{G}_x)^2, & T_{zz} &= n \Sigma (\bar{T}_z - \bar{G}_z)^2, \\ T_{xy} &= n \Sigma (\bar{T}_x - \bar{G}_x)(\bar{T}_y - \bar{G}_y), & T_{zy} &= n \Sigma (\bar{T}_z - \bar{G}_z)(\bar{T}_y - \bar{G}_y), \\ T_{yy} &= n \Sigma (\bar{T}_y - \bar{G}_y)^2, & T_{xz} &= n \Sigma (\bar{T}_x - \bar{G}_x)(\bar{T}_z - \bar{G}_z). \end{aligned}$$

To obtain the adjusted treatment sum of squares, one combines corresponding treatment and error variations and covariations, computes a multiple regression equation for the combined data, and computes the variation of the residuals about this equation. The adjusted treatment variation is obtained by subtracting the adjusted error variation from the variation of the residuals in the latter regression equation. The combined variations and covariations are as follows:

$$\begin{aligned} S_{xx} &= T_{xx} + E_{xx}, & S_{xz} &= T_{xz} + E_{xz}, & S_{xy} &= T_{xy} + E_{xy}, \\ S_{yy} &= T_{yy} + E_{yy}, & S_{zz} &= T_{zz} + E_{zz}, & S_{zy} &= T_{zy} + E_{zy}. \end{aligned}$$

The multiple regression equation associated with these combined data has the following form:

$$Y''_{ij} = b''_{y \cdot x}(X_{ij} - \bar{X}) + b''_{y \cdot z}(Z_{ij} - \bar{Z}) + \bar{Y}.$$

For the case of a single-factor experiment, this regression line is the one that would be obtained if the treatment classes were disregarded and an over-all

least-squares fit obtained for the entire set of experimental data. The regression coefficients are given by

$$b''_{y \cdot x} = \frac{S_{zz}S_{xy} - S_{xz}S_{zy}}{d''},$$

$$b''_{y \cdot z} = \frac{-S_{xz}S_{xy} - S_{zz}S_{zy}}{d''},$$

where $d'' = S_{xx}S_{zz} - S_{xz}^2$.

The sum of the squares of the residuals about this regression line is

$$S'_{yy} = (T + E)'_{yy} = S_{yy} - b''_{y \cdot x}S_{xy} - b''_{y \cdot z}S_{zy}.$$

The adjusted sum of squares for treatments is

$$T'_{yy} = (T + E)'_{yy} - E'_{yy}.$$

The analysis of covariance for this type of experiment is summarized in Table 11.7-1. Since two regression coefficients are estimated from the

Table 11.7-1 Analysis of Covariance, Two Covariates

Source	SS	df	MS	F
Treatments	T'_{yy}	$k - 1$	MS'_{treat}	$F = \frac{MS'_{\text{treat}}}{MS'_{\text{error}}}$
Error	E'_{yy}	$k(n - 1) - 2$	MS'_{error}	
Total	S'_{yy}	$kn - 3$		

experimental data to obtain the adjusted error, the degrees of freedom for the latter source of variation are $k(n - 1) - 2$.

To compute the variations and covariations, the procedures in parts i through iv of Table 11.3-1 may be extended to cover any number of covariates. For example,

$$E_{zz} = (2z) - (3z), \quad E_{xz} = (2xz) - (3xz),$$

$$T_{zz} = (3z) - (1z), \quad T_{xz} = (3xz) - (1xz).$$

In order to obtain the adjusted criterion variations, the pooled within-class as well as the over-all regression coefficients are required. Given the variations and covariations, the regression coefficients are obtained from the relations given in this section.

The adjusted mean for treatment j is

$$\bar{Y}'_j = \bar{Y}_j - b_{y \cdot x}(\bar{X}_j - \bar{X}) - b_{y \cdot z}(\bar{Z}_j - \bar{Z}).$$

The average effective error variance per experimental unit in making comparisons among adjusted treatment means is

$$s'^2_y = s'^2_{\text{error}} \left[1 + \frac{T_{xx}E_{zz} - 2T_{xz}E_{xz} + T_{zz}E_{xx}}{(k - 1)(E_{xx}E_{zz} - E_{xz}^2)} \right],$$

where $s_{\text{error}}'^2$ is the mean square for error in the analysis of covariance. The average effective error variance for the difference between two adjusted treatment means is

$$s_{\bar{Y}_j - \bar{Y}_k}'^2 = \frac{2s_y'^2}{n}.$$

The effective error variance for the difference between two adjusted treatment means for a test in the a priori category is

$$s_{\text{error}}'^2 \left[\frac{2}{n} + \frac{(\bar{X}_j - \bar{X}_k)^2 E_{zz} - 2(\bar{X}_j - \bar{X}_k)(\bar{Z}_j - \bar{Z}_k)E_{xz} + (\bar{Z}_j - \bar{Z}_k)^2 E_{xx}}{E_{xx}E_{zz} - E_{xz}^2} \right].$$

When the treatments affect the covariate, the last expression for the error variance is the more appropriate for use in testing differences between adjusted treatment means.

For the case of p covariates, expressions for the regression coefficients are most easily written in matrix notation. Suppose that the criterion is designated X_0 and the covariates X_1, X_2, \dots, X_p . Let

$$\begin{aligned} \mathbf{e}_0 &= [E_{01} \ E_{02} \ \cdots \ E_{0p}], & \mathbf{s}_0 &= [S_{01} \ S_{02} \ \cdots \ S_{0p}], \\ \mathbf{E} &= \begin{bmatrix} E_{11} & E_{12} & \cdots & E_{1p} \\ E_{21} & E_{22} & \cdots & E_{2p} \\ \vdots & \vdots & & \vdots \\ E_{p1} & E_{p2} & \cdots & E_{pp} \end{bmatrix}, & \mathbf{S} &= \begin{bmatrix} S_{11} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{22} & \cdots & S_{2p} \\ \vdots & \vdots & & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{pp} \end{bmatrix}, \\ \mathbf{b} &= [b_{0.1} \ b_{0.2} \ \cdots \ b_{0.p}], & \mathbf{b}_s &= [b_{0.1}'' \ b_{0.2}'' \ \cdots \ b_{0.p}'']. \end{aligned}$$

In terms of these matrices, the regression coefficients for the within-class data are

$$\mathbf{b} = \mathbf{e}_0 \mathbf{E}^{-1}.$$

The regression coefficients for the over-all data are

$$\mathbf{b}_s = \mathbf{s}_0 \mathbf{S}^{-1}.$$

The adjusted error and treatment variations are as follows:

$$E'_{yy} = E_{yy} - \mathbf{b} \mathbf{e}'_0, \quad T'_{yy} = S_{yy} - \mathbf{b}_s \mathbf{s}'_0 - E'_{yy}.$$

APPENDIX A

Topics Closely Related to the Analysis of Variance

A.1 Kruskal-Wallis H Test

Analogous to a single classification analysis of variance in which there are no repeated measures is the analysis of variance for ranked data. Suppose that there are k treatment classes having n_j observations in each class. Suppose further that the observations are in the form of ranks. That is, the criterion scores are ranks assigned irrespective of the treatment class to which an observation belongs. The data given below illustrate what is meant:

<i>Treatment 1</i>	<i>Treatment 2</i>	<i>Treatment 3</i>
1	3	6
2	5	9
4	8	12
7	10	13
	11	14

To test the hypothesis that the ranks within the treatment classes are a random sample from a common population of ranks, the following statistic may be used:

$$H = \frac{SS_{\text{treat}}}{MS_{\text{total}}}.$$

Numerator and denominator of the H statistic have the usual analysis-of-variance definitions.

When the hypothesis being tested is true, and when each n_j is larger than 5, the sampling distribution of this statistic may be approximated by a chi-square distribution having $k - 1$ degrees of freedom. For small values of n_j and k , special tables for the H statistic are available. Computational procedures for this test duplicate the procedures for a single classification analysis of variance. The latter procedures correct for tied ranks, whereas the specialized formulas for the H statistic require corrections for tied ranks, if these occur.

The Mann-Whitney U statistic is closely related to the H statistic when $k = 2$. Extensive tables for the U statistic for small n_j are available [see Siegel (1956, pp. 271–277)]. Individual comparisons between two treatments following an over-all H test may be made by means of the U statistic. An application of this procedure will be found in Lewis and Cotton (1958). If one of the treatments represents a control group, the nonparametric analogue of the Dunnett procedure is described in Sec. A.3 of this appendix.

A different approach for handling data which are in terms of ranks is to transform the ranks into normalized scores. Tables for making this transformation are given in Walker and Lev (1953, p. 480). In the latter form the data may be handled by means of the usual analysis of variance. The latter approach may lead to somewhat different conclusions. If the population to which inferences are to be made is considered to be one in which the criterion scores are normally distributed, then the analysis of variance in terms of the transformed scores is the more appropriate. On the other hand, if inferences are limited to ordinal measurement on the criterion scale, then the Kruskal-Wallis H statistic provides the more appropriate type of analysis.

A.2 Contingency Table with Repeated Measures

Consider an experiment in which n judges are asked to assign ranks to r products. Data obtained from this experiment may be summarized as follows:

Product	Rank						Total
	1	2	...	j	...	r	
1	n_{11}	n_{12}		n_{1j}		n_{1r}	n
2	n_{21}	n_{22}		n_{2j}		n_{2r}	n
.				.			.
.				.			.
.				.			.
i	n_{i1}	n_{i2}		n_{ij}		n_{ir}	n
.				.			.
.				.			.
.				.			.
r	n_{r1}	n_{r2}		n_{rj}		n_{rr}	n
Total	n	n		n		n	nr

In this summary, n_{ij} represents the number of times product i receives a rank of j . (The sampling distributions to be discussed in this section are obtained by limiting procedures which assume n to be large. The approximations have been shown to be reasonably close for $n = 30$ and larger. In similar limiting procedures, $n = 5$ and larger provide adequate approximations.)

To test the hypothesis of no differences between the products with respect to the frequency with which the products receive the rank of j , the statistic

$$Q_j = \frac{r \sum_i [n_{ij} - (n/r)]^2}{n} = \frac{(r \sum_i n_{ij}^2) - n^2}{n}$$

may be used. When there are no differences between the frequencies in column j , except those due to sampling error, Q_j has a sampling distribution which is approximated by a chi-square distribution having $r - 1$ degrees of freedom.

To test the over-all hypothesis of no differences between the ranks for the products, the statistic

$$Q = \sum Q_j$$

may be used. Under the hypothesis of no difference between the ranks assigned to the products, Anderson (1959) has shown that the statistic

$$\frac{(r-1)Q^2}{r} \doteq \chi^2 \quad \text{with df} = (r-1)^2.$$

If the statistic $(r-1)Q^2/r$ exceeds the critical value for a test having level of significance α , as determined from the appropriate sampling distribution, the hypothesis of no difference between the frequencies within the columns is rejected.

Anderson (1959) has shown that the Friedman statistic discussed in Sec. 4.7 provides a test on $r - 1$ components of the over-all chi square. The latter may be partitioned into individual comparisons, or contrasts, each having a single degree of freedom. In making tests on such comparisons, one uses the following estimates of the variances and covariances for the cell frequencies:

$$\text{var}(n_{ij}) = \frac{n(r-1)}{r^2},$$

$$\text{cov}(n_{ij}, n_{ik}) = \frac{-n}{r^2},$$

where n_{ij} and n_{ik} are two frequencies in the same row,

$$\text{cov}(n_{ij}, n_{kj}) = \frac{-n}{r^2},$$

where n_{ij} and n_{kj} are two frequencies in the same column, and

$$\text{cov}(n_{ij}, n_{km}) = \frac{n}{r^2(r-1)},$$

where n_{ij} and n_{km} are two frequencies in different rows and columns.

A 3×3 contingency table will be used for illustrative purposes. The cell frequencies are the column headings.

	n_{11}	n_{12}	n_{13}	n_{21}	n_{22}	n_{23}	n_{31}	n_{32}	n_{33}
C_1	0	0	0	-1	0	1	0	0	0
C_2	1	0	0	0	0	0	-1	0	0
C_3	1	0	-1	0	0	0	-1	0	1

The coefficients in row C_1 represent a linear comparison among the ranks assigned to product 2. The numerical value of the chi-square statistic corresponding to this comparison is

$$\chi_{C_1}^2 = \frac{(n_{23} - n_{21})^2}{\text{var}(n_{23} - n_{21})}.$$

The denominator of this statistic is

$$\begin{aligned} \text{var}(n_{23} - n_{21}) &= \text{var}(n_{23}) + \text{var}(n_{21}) - 2 \text{cov}(n_{23}, n_{21}) \\ &= \frac{n(r-1)}{r^2} + \frac{n(r-1)}{r^2} - \frac{2(-n)}{r^2} \\ &= \frac{2n}{r}. \end{aligned}$$

The above chi-square statistic has one degree of freedom. Should this statistic exceed the critical value for an α -level test, the data would indicate a statistically significant difference between the rankings assigned to product 2.

The coefficients in row C_2 represent a linear comparison among the products for rank 1. (This comparison is not orthogonal to C_1 .) The chi-square statistic corresponding to this comparison is

$$\begin{aligned} \chi_{C_2}^2 &= \frac{(n_{11} - n_{31})^2}{\text{var}(n_{11} - n_{31})} = \frac{(n_{11} - n_{31})^2}{2n/r} \\ &= \frac{r(n_{11} - n_{31})^2}{2n}. \end{aligned}$$

This chi-square statistic has one degree of freedom.

The coefficients in row C_3 represent a comparison between the differences in linear rankings for products 1 and 3. (Comparison C_3 is orthogonal to comparison C_1 .) The chi-square statistic corresponding to this comparison is

$$\chi_{C_3}^2 = \frac{(n_{11} - n_{13} + n_{33} - n_{31})^2}{\text{var}(n_{11} - n_{13} + n_{33} - n_{31})}.$$

Table A.2-1 Numerical Example

(i)

Judge	Product			Total
	<i>a</i>	<i>b</i>	<i>c</i>	
1	2	1	3	6 = P_1
2	1	2	3	6
3	1	3	2	6
4	1	2	3	6
5	2	1	3	6
6	1	2	3	6
7	1	3	2	6
8	1	2	3	6
$T_a = 10$	16	22		48

$$G^2/nr = (48^2)/8(3) = 96.00$$

$$(\sum T_j^2)/n = 105.00$$

$$\sum X^2 = 112$$

$$(\sum P_i^2)/r = 96.00$$

$$SS_{\text{products}} = 105 - 96.00 = 9.00$$

$$SS_{\text{w. judges}} = 112 - 96.00 = 16.00$$

$$MS_{\text{w. judges}} = 16.00/16 = 1.00$$

$$\chi_{\text{ranks}}^2 = \frac{SS_{\text{products}}}{MS_{\text{w. judges}}} = 9.00$$

(ii)

Product	Rank			Total
	1	2	3	
<i>a</i>	6	2	0	8
<i>b</i>	2	4	2	8
<i>c</i>	0	2	6	8
	8	8	8	

$$\chi_{C_1}^2 = 3(2 - 2)^2/2(8) = 0$$

$$\chi_{C_3}^2 = 2(6 - 0 + 6 - 0)/4(8) = 9$$

$$\chi_{C_1}^2 + \chi_{C_3}^2 = 9$$

The individual variances and covariances required to obtain the term in the denominator are given by

$$\begin{aligned} \text{var}(n_{11} - n_{13} + n_{33} - n_{31}) &= 4 \text{var}(n_{ij}) - 4 \text{cov}(n_{ij}, n_{ik}) \\ &\quad - 4 \text{cov}(n_{ij}, n_{kj}) + 4 \text{cov}(n_{ij}, n_{km}) \\ &= \frac{4n}{r-1} \end{aligned}$$

Thus,

$$\chi_{C_3}^2 = \frac{(r-1)(n_{11} - n_{13} + n_{33} - n_{31})^2}{4n}$$

The chi-square statistic used in the Friedman test is equivalent to the sum of $r-1$ orthogonal comparisons among the products. For the

case of a 3×3 contingency table, C_1 and C_3 are orthogonal comparisons of this kind. Hence,

$$\chi_{C_1}^2 + \chi_{C_3}^2 = \chi_{\text{ranks}}^2 = \frac{SS_{\text{products}}}{MS_{\text{w. judge}}},$$

where the last term on the right is the statistic used in the Friedman test.

The numerical example given in Table A.2-1 illustrates this last relationship. Basic data are given in part i. There are $n = 8$ judges and $r = 3$ products. The rankings assigned by each judge are shown. The variations obtained in part i are defined as follows:

$$SS_{\text{products}} = \frac{\sum T_j^2}{n} - \frac{G^2}{nr},$$

$$SS_{\text{w. judge}} = \sum X^2 - \frac{\sum P_i^2}{r}.$$

Computation of this latter source of variation may be simplified when no tied ranks are permitted. The critical value for a .05-level test is $\chi_{.95}^2(2) = 6.00$. Hence the test in part i indicates that the differences in ranks assigned to the products are statistically significant.

Data from part i are rearranged to form a contingency table in part ii. The comparison C_1 , which was defined earlier in this section, indicates no difference in the linear ranking for product b . The critical value associated with a .05-level test on C_3 is $\chi_{.95}^2(1) = 3.8$. Hence the data indicate that there is a statistically significant difference between the linear rankings for products a and c . Note that

$$\chi_{C_1}^2 + \chi_{C_3}^2 = \chi_{\text{ranks}}^2.$$

A.3 Comparing Treatment Effects with a Control

Procedures for comparing all treatments with a control were discussed in Sec. 3.10. A nonparametric analogue of these procedures has been developed by Steel (1959). A numerical example will be used to illustrate the procedures for comparing all treatments with a control when data are in terms of ranks. In a sense, these comparisons are part of the over-all hypothesis tested by the Kruskal-Wallis H statistic.

The basic data for this numerical example are given in part i of Table A.3-1. Suppose that only the rank order of these measurements is considered meaningful. The data in part ii are in terms of ranks. To obtain these ranks, the control scores and the treatment a scores are combined; then ranks 1 to $2n$ are assigned to the combined set of scores. In case of ties, the mean of the tied ranks is used. The combined sets of scores for the control and treatment a groups are as follows:

Scores . . .	35	40	45	<u>45</u>	48	50	<u>50</u>	<u>60</u>	<u>62</u>	<u>75</u>
Ranks . . .	1	2	3.5	<u>3.5</u>	5	6.5	<u>6.5</u>	<u>8</u>	<u>9</u>	10

Data from the control condition are underscored. The combined sets of scores from the control and treatment c conditions are as follows:

Scores . . .	<u>45</u>	<u>50</u>	<u>60</u>	<u>62</u>	<u>75</u>	75	78	80	80	84
Ranks . . .	<u>1</u>	<u>2</u>	<u>3</u>	<u>4</u>	<u>5.5</u>	5.5	7	8.5	8.5	10

The sum of the ranks for the control group and each of the treatment groups is then computed. T'_a represents the sum of ranks for the control condition when the scores are ranked with reference to treatment a . The

Table A.3-1 Numerical Example

	Control			Treatment a	Treatment b	Treatment c
	a	b	c			
(i)	45			35	58	75
	50			40	62	78
	60	$n = 5$		45	70	80
	62			48	78	80
	75			50	80	84
(ii)	3.5	1	1	1	3	5.5
	6.5	2	2	2	6	7
	8	4	3	3.5	8	8.5
	9	5	4	5	9	8.5
	10	7	5.5	6.5	10	10
	<u>37.0</u>	<u>19.0</u>	<u>15.5</u>	<u>18.0</u>	<u>36.0</u>	<u>39.5</u>
	T'_a	T'_b	T'_c	T_a	T_b	T_c

test statistic used in the decision rule about the difference between treatment a and the control condition is $\min(T'_a, T_a)$, that is, the smaller of T'_a and T_a . In this case,

$$\min(T'_a, T_a) = \min(37, 18) = 18.$$

As a partial check on the numerical work,

$$T'_i + T_i = n(2n + 1).$$

Steel (1959) has constructed tables of the sampling distribution of the statistic $\min(T'_i, T_i)$. Probabilities in these tables are in terms of an experimentwise error rate. By definition, the latter is the ratio of the number of experiments with one or more false significance statements to the total number of experiments. For the case $n = 5$ and $k = 3$, where k is the number of treatments (excluding the control), the critical value for the rank sum statistic for a two-tailed test with error rate .05 is 16. The decision is made to reject the hypothesis of no difference between treatment i and the control if

$$\min(T'_i, T_i) \leq 16.$$

For the data in Table A.3-1, treatment c is statistically different from the control, but none of the other differences between the treatments and the control is statistically significant, with a .05-level experimentwise error rate. Had the direction of the differences between the control and the experimental groups been predicted prior to the experiment, one-tailed rather than two-tailed tests would be appropriate. The critical value for a .05-level one-tailed test in which $n = 5$ and $k = 3$ is 18.

A.4 General Partition of Degrees of Freedom in a Contingency Table

To illustrate the procedures to be discussed in this section, consider the three-dimensional contingency table having the following form (all observations are assumed to be independent):

	c_1		c_2	
	b_1	b_2	b_1	b_2
a_1	n_{111}	n_{121}	n_{112}	n_{122}
a_2	n_{211}	n_{221}	n_{212}	n_{222}
a_3	n_{311}	n_{321}	n_{312}	n_{322}

In general there will be p classes for category A , q classes for category B , and r classes for category C . The frequency in cell abc_{ijk} will be designated by the symbol n_{ijk} .

If the B category in the above contingency table is disregarded, the resulting AC summary table will have the following form.

	c_1	c_2	Total
a_1	$n_{1..}$	$n_{1..}$	$n_{1..}$
a_2	$n_{2..}$	$n_{2..}$	$n_{2..}$
a_3	$n_{3..}$	$n_{3..}$	$n_{3..}$
	$n_{..1}$	$n_{..2}$	$n_{...}$

In general the following notation will be used:

$$\begin{aligned} \sum_i n_{ijk} &= n_{.jk}, & \sum_j n_{ijk} &= n_{i.k}, & \sum_k n_{ijk} &= n_{ij.}; \\ \sum_i \sum_j n_{ijk} &= \sum_i n_{i.k} = n_{..k}, & \sum_i \sum_k n_{ijk} &= \sum_i n_{ij.} = n_{.j.}; \\ \sum_i \sum_j \sum_k n_{ijk} &= \sum_i \sum_j n_{ij.} = \sum_i n_{i..} = n_{...} \end{aligned}$$

If sampling is random with respect to all categories, $pqr - 1$ parameters are necessary to specify the population from which the sample of size $n_{...}$ was drawn. These parameters may be specified in terms of the following proportions:

P_{ijk} = proportion of population frequency in cell abc_{ijk} .

The expected frequency in cell abc_{ijk} , which will be designated by the symbol n'_{ijk} , is

$$n'_{ijk} = P_{ijk}n_{...}$$

The expected frequencies for the marginal totals of category A would be

$$n'_{i..} = \sum_j \sum_k n'_{ijk};$$

alternatively,

$$n'_{i..} = \sum_j \sum_k P_{ijk}n_{...} = P_{i..}n_{...}$$

The symbol $P_{i..}$ designates the population proportion for the category a_i . The other expected marginal frequencies are

$$n'_{.j.} = P_{.j.}n_{...},$$

$$n'_{..k} = P_{..k}n_{...}.$$

The expected frequency for a cell in the AB summary table is given by

$$\begin{aligned} n'_{ij.} &= \sum_k n'_{ijk} \\ &= \sum_k P_{ijk}n_{...} = P_{ij.}n_{...} \end{aligned}$$

Other expected frequencies for two-way summary tables are

$$n'_{i.k} = P_{i.k}n_{...},$$

$$n'_{.jk} = P_{.jk}n_{...}.$$

If all of the $pqr - 1$ parameters in the population are specified by an a priori model, and if the sampling is random with respect to all categories, then the total chi square indicated in Table A.4-1 may be partitioned in the manner shown in this table. This partition bears a marked resemblance to an analysis-of-variance table.

Tests with respect to conformity with the specified model may be made, provided that the sampling distributions for the statistics indicated may be approximated by chi-square distributions. If each of the expected cell frequencies is greater than 5, the chi-square distributions will provide good approximations. If a relatively small number of expected frequencies are less than 5, the chi-square approximations will still be good.

A review of some of the work that has been done on the partition of chi square in contingency tables will be found in Sutcliffe (1957). If the model for the population can be completely specified on a priori grounds, and if the sampling is random with respect to all categories, then the method of partition indicated in Table A.4-1 may be carried out quite readily. In practice, however, certain of the parameters in the model are often estimated from the observed data. For example, the parameters $P_{i..}$, $P_{.j.}$, and $P_{..k}$ may be estimated from the marginal frequencies of the sample data. Under the

hypothesis of no interactions of any order (i.e., no two-category or no three-category interactions), the expected proportion for cell abc_{ijk} is

$$P_{ijk} = P_{i..}P_{.j.}P_{..i},$$

and the expected frequency in cell abc_{ijk} is

$$n'_{ijk} = P_{ijk}n_{...}$$

Under this method for specifying the model for the population, the total

Table A.4-1 Partition of Chi Square

Source	Chi square	df
Total	$\chi^2_{\text{total}} = \Sigma\Sigma\Sigma[(n_{ijk} - n'_{ijk})^2/n'_{ijk}]$	$pqr - 1$
A	$\chi^2_a = \Sigma[(n_{i..} - n'_{i..})^2/n'_{i..}]$	$p - 1$
B	$\chi^2_b = \Sigma[(n_{.j.} - n'_{.j.})^2/n'_{.j.}]$	$q - 1$
C	$\chi^2_c = \Sigma[(n_{..k} - n'_{..k})^2/n'_{..k}]$	$r - 1$
AB	$\chi^2_{ab} = \Sigma\Sigma[(n_{ij.} - n'_{ij.})^2/n'_{ij.}] - \chi^2_a - \chi^2_b$	$(p - 1)(q - 1)$
AC	$\chi^2_{ac} = \Sigma\Sigma[(n_{i.k} - n'_{i.k})^2/n'_{i.k}] - \chi^2_a - \chi^2_c$	$(p - 1)(r - 1)$
BC	$\chi^2_{bc} = \Sigma\Sigma[(n_{.jk} - n'_{.jk})^2/n'_{.jk}] - \chi^2_b - \chi^2_c$	$(q - 1)(r - 1)$
ABC	$\chi^2_{abc} = \chi^2_{\text{total}} - \chi^2_a - \chi^2_b - \chi^2_c - \chi^2_{ab} - \chi^2_{ac} - \chi^2_{bc}$	$(p - 1)(q - 1)(r - 1)$

chi square may be partitioned as shown in Table A.4-2. In this case, note that the degrees of freedom for the total chi square are

$$(pqr - 1) - (p - 1) - (q - 1) - (r - 1) = pqr - p - q - r + 2.$$

Should the three-factor interaction be statistically significant in this type of analysis, the two-way summary tables should be studied separately within

Table A.4-2 Partition of Chi Square When Probabilities are Estimated from Marginal Totals

Source	Chi square	df
Total	$\chi^2_{\text{total}} = \Sigma\Sigma\Sigma[(n_{ijk} - n'_{ijk})^2/n'_{ijk}]$	$(pqr - 1) - (p - 1) - (q - 1) - (r - 1)$
AB	$\chi^2_{ab} = \Sigma\Sigma[(n_{ij.} - n'_{ij.})^2/n'_{ij.}]$	$(p - 1)(q - 1)$
AC	$\chi^2_{ac} = \Sigma\Sigma[(n_{i.k} - n'_{i.k})^2/n'_{i.k}]$	$(p - 1)(r - 1)$
BC	$\chi^2_{bc} = \Sigma\Sigma[(n_{.jk} - n'_{.jk})^2/n'_{.jk}]$	$(q - 1)(r - 1)$
ABC	$\chi^2_{abc} = \chi^2_{\text{total}} - \chi^2_{ab} - \chi^2_{ac} - \chi^2_{bc}$	$(p - 1)(q - 1)(r - 1)$

a fixed level of the third category. In these latter tables, the marginal totals may be used in some cases to estimate the cell frequencies. For example, if the AB data for level c_1 are being studied, under the hypothesis of no interaction between categories A and B for level c_1 ,

$$n'_{ij1} = \frac{n_{i.1}n_{.j1}}{n_{..1}}.$$

This expected value for cell $ij1$ will not in general be the same as that obtained under the hypothesis of no interactions of any order.

Another case which arises in practice is one in which the sampling is restricted with respect to the number of observations in each of the cells of the form ab_{ij} but random with respect to the category C . If the marginal totals are used in the estimation of $P_{..k}$, then

$$P_{..k} = \frac{n_{..k}}{n_{...}}, \quad P_{ij.} = \frac{n_{ij.}}{n_{...}}.$$

Under the hypothesis of no interactions,

$$n'_{ijk} = P_{ij.} P_{..k} n_{...}.$$

In this case, chi square may be partitioned in the following manner:

Source	df
Total	$(pq - 1)(r - 1)$
AC	$(p - 1)(r - 1)$
BC	$(q - 1)(r - 1)$
ABC	$(p - 1)(q - 1)(r - 1)$

Should the three-factor interaction prove to be statistically significant in this case, it is advisable to study the equivalent of simple effects for category C at each of the separate levels of factors A and B .

A.5 Hotelling's T^2 Test for the Equality of k Means

In the chapters dealing with repeated measures, implicit in the final F tests on within-subject effects was the assumption that the variance-covariance matrix had symmetry of the following form:

$$(1) \quad \Sigma = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix}.$$

Over-all tests on within-subject effects can be made through use of Hotelling's T^2 statistic without the assumption of symmetry of the above form. The only assumptions underlying the latter tests are that the variables have a multivariate normal distribution with variance-covariance matrix Σ , where the general form of Σ is arbitrary.

For a single-classification repeated-measure experiment in which there are k treatments and n subjects, Hotelling's T^2 has the form

$$T^2 = n\bar{\mathbf{x}}'S^{-1}\bar{\mathbf{x}},$$

where

$$\bar{\mathbf{x}} = \begin{bmatrix} \bar{T}_1 - \bar{G} \\ \bar{T}_2 - \bar{G} \\ \vdots \\ \bar{T}_k - \bar{G} \end{bmatrix},$$

and S^{-1} is the inverse of the variance-covariance matrix. The hypothesis that $\sigma_r^2 = 0$ is rejected if

$$T_{\text{obs}}^2 > \frac{(n-1)k}{n-k} F_{1-\alpha}(k, n-k).$$

To illustrate the computation of Hotelling's T^2 statistic, suppose that the following data have been obtained in an experiment in which $n = 15$ and $k = 3$:

$$\begin{aligned} \bar{T}_1 &= 10, & \bar{T}_2 &= 14, & \bar{T}_3 &= 18, & \bar{G} &= 14; \\ S &= \begin{bmatrix} 100.00 & 40.00 & 0.00 \\ 40.00 & 50.00 & 40.00 \\ 0.00 & 40.00 & 50.00 \end{bmatrix}, & \text{var} - \text{cov} &= 40.00. \end{aligned}$$

From these data on the treatment means,

$$SS_{\text{treat}} = n\Sigma(\bar{T}_j - \bar{G})^2 = 480.00,$$

$$MS_{\text{treat}} = \frac{480.00}{2} = 240.00.$$

Hence the analysis of variance for the within-subject effects is as follows:

Source	df	MS	F
Within subjects	30		
Treatments	2	240.00	6.00
Residual	28	40.00	

The critical value for a .01-level test, under the assumption that Σ has the symmetry indicated by (1), is

$$F_{.99}[k-1, (n-1)(k-1)] = F_{.99}(2, 28) = 5.45.$$

By using the conservative test suggested by Box, the approximate critical value for a .01-level test, without the symmetry assumption on Σ , is

$$F_{.99}(1, n-1) = F_{.99}(1, 14) = 8.86.$$

Hotelling's T^2 statistic is given by

$$T_{\text{obs}}^2 = 15 \begin{bmatrix} -4.00 & 0.00 & 4.00 \end{bmatrix} \begin{bmatrix} .090 & -.200 & .160 \\ -.200 & .500 & -.400 \\ .160 & -.400 & .500 \end{bmatrix} \begin{bmatrix} -4.00 \\ 0.00 \\ 4.00 \end{bmatrix} \\ = 15(2.08) = 31.20.$$

The matrix S^{-1} is obtained from the matrix S by the Dwyer square-root method or by the Doolittle method.

The critical value for a .01-level test using the T^2 statistic is

$$\frac{(14)(3)}{12} F_{.99}(3, 12) = 22.31.$$

In this case the T^2 test leads to the rejection of the hypothesis that $\sigma_{\tau}^2 = 0$, whereas the approximate test does not lead to rejection of this hypothesis.

Table A.5-1 Numerical Example (Repeated Measures)

($k = 4, n = 7$)

Observed means: $\bar{T}_1 = 16, \bar{T}_2 = 22, \bar{T}_3 = 22, \bar{T}_4 = 20, \bar{G} = 20$

Estimate of Σ (assuming symmetry) obtained from observed data:

$$S = \begin{bmatrix} 3.00 & 1.80 & 1.80 & 1.80 \\ 1.80 & 3.00 & 1.80 & 1.80 \\ 1.80 & 1.80 & 3.00 & 1.80 \\ 1.80 & 1.80 & 1.80 & 3.00 \end{bmatrix}$$

(i)

$$S^{-1} = \begin{bmatrix} .654 & -.178 & -.178 & -.178 \\ -.178 & .654 & -.178 & -.178 \\ -.178 & -.178 & .654 & -.178 \\ -.178 & -.178 & -.178 & .654 \end{bmatrix}$$

$$MS_{\text{treat}} = \frac{n \sum (\bar{T}_j - \bar{G})^2}{k - 1} = \frac{7[(-4)^2 + 2^2 + 2^2 + 0^2]}{3} \\ = 56.00$$

$$(ii) \quad MS_{\text{res}} = \overline{\text{var}} - \overline{\text{cov}} = 3.00 - 1.80 = 1.20$$

$$F = \frac{MS_{\text{treat}}}{MS_{\text{res}}} = \frac{56.00}{1.20} = 46.67$$

$$(iii) \quad T^2 = 7 \begin{bmatrix} -4 & 2 & 2 & 0 \end{bmatrix} S^{-1} \begin{bmatrix} -4 & 2 & 2 & 0 \end{bmatrix}' = 140.0$$

Instead of a single-classification experiment, suppose that one had the factorial experiment:

	b_1	b_2	...	b_q
a_1	G_1	G_1		G_1
\vdots				
a_p	G_p	G_p		G_p

In using the T^2 statistic for testing effects associated with factor B , the following homogeneity assumptions are required:

$$\Sigma_{a_1} = \Sigma_{a_2} = \cdots = \Sigma_{a_p} = \Sigma.$$

In words, the variance-covariance matrices for each level of factor A must be equal to Σ .

When the matrix Σ has the symmetry indicated in equation (1), it is relatively easy to show numerically that Hotelling's T^2 is proportional to the usual F statistic in the analysis of variance. The data in Table A.5-1 will be used for this purpose.

In part i the observed treatment means and an estimate, S , of the matrix Σ , assuming symmetry of the type indicated in equation (1), are given. From S one computes S^{-1} . The latter has the same kind of symmetry that S has. The F ratio that would be obtained from the usual analysis of variance is computed in part ii. The T^2 statistic is computed in part iii. It is noted that

$$\begin{aligned} T_{\text{obs}}^2 &= (k - 1)F_{\text{obs}} \\ &= (4 - 1)(46.67) = 140.0. \end{aligned}$$

A.6 Least-squares Estimators—General Principles

A simple example will be used to illustrate the least-squares principle in estimation. These principles will then be stated in general form. Suppose that the observations x_1 , x_2 , and x_3 have the following structural form:

$$(1) \quad \begin{aligned} x_1 &= \beta_1 + \varepsilon_1, \\ x_2 &= \beta_1 + 2\beta_2 + \varepsilon_2, \\ x_3 &= \beta_2 + \varepsilon_3. \end{aligned}$$

In (1), β_1 and β_2 are unknown parameters; however, the coefficients of β_1 and β_2 are assumed to be known. The variates ε_1 , ε_2 , and ε_3 are assumed to be independently and normally distributed, with expected value zero and constant variance σ_ε^2 . The problem is to find linear estimates of β_1 , β_2 , and β_3 such that the estimate of σ_ε^2 is minimized. Let b_1 , b_2 , and b_3 be estimates of the corresponding β 's.

To obtain these estimates, one minimizes the expression

$$(2) \quad (x_1 - b_1)^2 + (x_2 - b_1 - 2b_2)^2 + (x_3 - b_2)^2.$$

Differentiating (2) with respect to b_1 and b_2 , setting the resulting expressions equal to zero, and then rearranging the terms yields the following normal equations:

$$(3) \quad \begin{aligned} 2b_1 + 2b_2 &= x_1 + x_2, \\ 2b_1 + 5b_2 &= 2x_2 + x_3. \end{aligned}$$

The problem now is to solve (3) for b_1 and b_2 . It is noted that (3) is linear in the x 's as well as linear in the b 's. If further the set (3) is *linearly independent* (in this case they are), then one may obtain unique values for the

b 's as linear functions of the x 's. (A set of equations is linearly independent if no equation in the set is a multiple of any other or a weighted sum of two or more of the other equations.)

Solving (3), one obtains

$$(4) \quad \begin{aligned} b_1 &= \frac{1}{6}(5x_1 + x_2 - 2x_3), \\ b_2 &= -\frac{1}{3}(x_1 - x_2 - x_3). \end{aligned}$$

Thus b_1 and b_2 are linear functions of the x 's (that is, a weighted sum of the x 's). The variance of any linear function of the x 's is a linear function of the corresponding variances. In this case,

$$\begin{aligned} \text{var}(b_1) &= \frac{1}{36}[(5)^2 + (1)^2 + (-2)^2]\sigma_\varepsilon^2, \\ \text{var}(b_2) &= \frac{1}{9}[(1)^2 + (-1)^2 + (-1)^2]\sigma_\varepsilon^2. \end{aligned}$$

The coefficient of σ_ε^2 is the sum of the squares of the coefficients of the x 's in the corresponding b 's. The variance of each of the x 's is assumed to be σ_ε^2 . The smaller the variance of the b 's, the "better" the estimate. The best linear unbiased estimator of a b is the one having the smallest variance. In this case,

$$\begin{aligned} \text{var}(b_1) &= \frac{30}{36}\sigma_\varepsilon^2 = \frac{5}{6}\sigma_\varepsilon^2, \\ \text{var}(b_2) &= \frac{3}{9}\sigma_\varepsilon^2 = \frac{1}{3}\sigma_\varepsilon^2. \end{aligned}$$

The least-squares estimators b_1 and b_2 can be shown to be the best linear unbiased estimators of β_1 and β_2 . The proof of the general case is given in a basic theorem in statistics—the Markoff theorem. By way of contrast, let

$$b'_1 = x_2 - 2x_3.$$

$$\begin{aligned} \text{From (1),} \quad x_2 - 2x_3 &= \beta_1 + 2\beta_2 + \varepsilon_2 - 2\beta_2 - 2\varepsilon_3 \\ &= \beta_1 + \varepsilon_2 - 2\varepsilon_3. \end{aligned}$$

Hence b'_1 is a linear unbiased estimate of β_1 . The variance of b'_1 is

$$\text{var}(b'_1) = [(1)^2 + (-2)^2]\sigma_\varepsilon^2 = 5\sigma_\varepsilon^2.$$

The corresponding least-squares estimate b_1 has variance $\frac{5}{6}\sigma_\varepsilon^2$.

To illustrate (1) through (4) numerically, suppose that one has the following experimentally determined values for the x 's:

$$x_1 = 6, \quad x_2 = 0, \quad x_3 = -6.$$

From (4) the least-squares estimators are

$$\begin{aligned} b_1 &= \frac{1}{6}[5(6) + (0) - 2(-6)] = 7, \\ b_2 &= -\frac{1}{3}[(6) - (0) - (-6)] = -4. \end{aligned}$$

Substituting these values for the x 's and the b 's in (2), one obtains an estimate of the sum of squares due to the ε 's.

$$\begin{aligned} \text{SS}_{\text{error}} &= (6 - 7)^2 + (0 - 7 + 8)^2 + (-6 + 4)^2 \\ &= 6. \end{aligned}$$

An algebraically equivalent method for computing SS_{error} is given by

$$(5) \quad SS_{\text{error}} = \sum x^2 - b_1 f_1 - b_2 f_2,$$

where f_1 is the linear function of the x 's on the right-hand side of the first equation in (3) and f_2 is the corresponding linear function in the second equation in (3). In this case

$$\begin{aligned} SS_{\text{error}} &= [(6)^2 + (0)^2 + (-6)^2] - [(7)(6 + 0)] - [(-4)(2 \cdot 0 - 6)] \\ &= 72 - 42 - 24 = 6. \end{aligned}$$

For the general case, suppose that the set (1a) consists of equations of the following form,

$$(1a) \quad \begin{aligned} x_1 &= a_{11}\beta_1 + a_{12}\beta_2 + \cdots + a_{1k}\beta_k + \varepsilon_1, \\ &\dots\dots\dots, \\ x_n &= a_{n1}\beta_1 + a_{n2}\beta_2 + \cdots + a_{nk}\beta_k + \varepsilon_n, \end{aligned}$$

where $n > k$. The β 's are unknown parameters, the a 's are known constants, and the ε 's are independently and normally distributed, with mean zero and constant variance σ_ε^2 .

The normal equations will have the general form

$$(3a) \quad \begin{aligned} g_1(b_1, \dots, b_k) &= f_1(x_1, \dots, x_n), \\ &\dots\dots\dots, \\ g_k(b_1, \dots, b_k) &= f_k(x_1, \dots, x_n), \end{aligned}$$

where the g 's and f 's denote linear functions. These equations may or may not be linearly independent. If they are not, a set of linear restrictions may be placed on the b 's so as to yield a unique set of values. These restrictions generally have a form which tends to simplify the solution of the resulting equations. Although these side conditions do not influence tests on differences between effects or contrasts in general, they do influence the estimates of the effects per se. Hence such side conditions should not be arbitrary in form; rather, they should reflect the structure of the population about which inferences are to be drawn, if the estimates per se are used for descriptive purposes. The side conditions are generally of the form

$$u_1 b_1 + u_2 b_2 + \cdots + u_k b_k = 0,$$

i.e., the side conditions do not involve the x 's.

If one solves system (3a) plus the side conditions for the b 's, one obtains expressions of the following form:

$$(4a) \quad \begin{aligned} b_1 &= c_{11}f_1 + c_{12}f_2 + \cdots + c_{1k}f_k, \\ &\dots\dots\dots, \\ b_k &= c_{k1}f_1 + c_{k2}f_2 + \cdots + c_{kk}f_k. \end{aligned}$$

That is, the b 's are linear functions of the f 's; the latter in turn are linear functions of the x 's. Hence the general form of b_j is

$$b_j = d_{j1}x_1 + d_{j2}x_2 + \cdots + d_{jn}x_n.$$

Here b_j is a linear unbiased estimate of β_j . Further,

$$\text{var}(b_j) = (d_{j1}^2 + d_{j2}^2 + \cdots + d_{jn}^2)\sigma_\epsilon^2.$$

Of all linear unbiased estimators of β_j , the least-squares estimator b_j can be shown to have minimum variance. Hence b_j is the best linear unbiased estimator which satisfies the set (3a) and the side conditions (if any).

An estimate of σ_ϵ^2 is given by

$$\frac{\text{SS}_{\text{error}}}{\text{df}_{\text{error}}} = \frac{\Sigma x^2 - \Sigma b_i f_i}{\text{df}_{\text{error}}},$$

where $\text{df}_{\text{error}} = n - (\text{number of linearly independent parameters estimated})$.

APPENDIX B

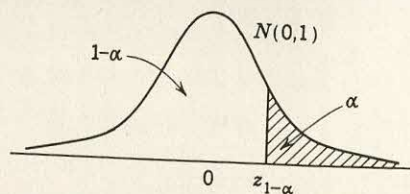
Tables

- B.1 Unit normal distribution
- B.2 Student's t distribution
- B.3 F distribution
- B.4 Distribution of the studentized range statistic
- B.5 Arcsin transformation
- B.6 Distribution of t statistic in comparing treatment means with a control
- B.7 Distribution of F_{\max} statistic
- B.8 Critical values for Cochran's test for homogeneity of variance
- B.9 Chi-square distribution
- B.10 Coefficients of orthogonal polynomials
- B.11 Curves of constant power for tests on main effects
- B.12 Random permutations of 16 numbers

Table B.1 Unit Normal Distribution*

$$[P(z \leq z_{1-\alpha}) = 1 - \alpha]$$

$1 - \alpha$	$z_{1-\alpha}$	$1 - \alpha$	$z_{1-\alpha}$	$1 - \alpha$	$z_{1-\alpha}$
.50	0.00	.75	0.67	.950	1.645
.51	0.03	.76	0.71	.955	1.695
.52	0.05	.77	0.74	.960	1.751
.53	0.08	.78	0.77	.965	1.812
.54	0.10	.79	0.81	.970	1.881
.55	0.13	.80	0.84	.975	1.960
.56	0.15	.81	0.88	.980	2.054
.57	0.18	.82	0.92	.985	2.170
.58	0.20	.83	0.95	.990	2.326
.59	0.23	.84	0.99	.995	2.576
.60	0.25	.85	1.04	.996	2.652
.61	0.28	.86	1.08	.997	2.748
.62	0.30	.87	1.13	.998	2.878
.63	0.33	.88	1.17	.999	3.090
.64	0.36	.89	1.23		
.65	0.39	.90	1.28	.9995	3.291
.66	0.41	.91	1.34	.9995	3.891
.67	0.44	.92	1.41		
.68	0.47	.93	1.48	.99995	4.417
.69	0.50	.94	1.55		
.70	0.52			.999995	5.327
.71	0.55				
.72	0.58				
.73	0.61				
.74	0.64				



* This table is abridged from Table 9 in *Biometrika Tables for Statisticians*, vol. 1. (2d ed.) New York: Cambridge, 1958. Edited by E. S. Pearson and H. O. Hartley. Reproduced with the kind permission of E. S. Pearson and the trustees of *Biometrika*.

Table B.2 Student's t Distribution

df	Percentile point						
	70	80	90	95	97.5	99	99.5
1	.73	1.38	3.08	6.31	12.71	31.82	63.66
2	.62	1.06	1.89	2.92	4.30	6.96	9.92
3	.58	.98	1.64	2.35	3.18	4.54	5.84
4	.57	.94	1.53	2.13	2.78	3.75	4.60
5	.56	.92	1.48	2.01	2.57	3.36	4.03
6	.55	.91	1.44	1.94	2.45	3.14	3.71
7	.55	.90	1.42	1.90	2.36	3.00	3.50
8	.55	.89	1.40	1.86	2.31	2.90	3.36
9	.54	.88	1.38	1.83	2.26	2.82	3.25
10	.54	.88	1.37	1.81	2.23	2.76	3.17
11	.54	.88	1.36	1.80	2.20	2.72	3.11
12	.54	.87	1.36	1.78	2.18	2.68	3.06
13	.54	.87	1.35	1.77	2.16	2.65	3.01
14	.54	.87	1.34	1.76	2.14	2.62	2.98
15	.54	.87	1.34	1.75	2.13	2.60	2.95
16	.54	.86	1.34	1.75	2.12	2.58	2.92
17	.53	.86	1.33	1.74	2.11	2.57	2.90
18	.53	.86	1.33	1.73	2.10	2.55	2.88
19	.53	.86	1.33	1.73	2.09	2.54	2.86
20	.53	.86	1.32	1.72	2.09	2.53	2.84
21	.53	.86	1.32	1.72	2.08	2.52	2.83
22	.53	.86	1.32	1.72	2.07	2.51	2.82
23	.53	.86	1.32	1.71	2.07	2.50	2.81
24	.53	.86	1.32	1.71	2.06	2.49	2.80
25	.53	.86	1.32	1.71	2.06	2.48	2.79
26	.53	.86	1.32	1.71	2.06	2.48	2.78
27	.53	.86	1.31	1.70	2.05	2.47	2.77
28	.53	.86	1.31	1.70	2.05	2.47	2.76
29	.53	.85	1.31	1.70	2.04	2.46	2.76
30	.53	.85	1.31	1.70	2.04	2.46	2.75
40	.53	.85	1.30	1.68	2.02	2.42	2.70
50	.53	.85	1.30	1.67	2.01	2.40	2.68
60	.53	.85	1.30	1.67	2.00	2.39	2.66
80	.53	.85	1.29	1.66	1.99	2.37	2.64
100	.53	.84	1.29	1.66	1.98	2.36	2.63
200	.52	.84	1.29	1.65	1.97	2.34	2.60
500	.52	.84	1.28	1.65	1.96	2.33	2.59
∞	.52	.84	1.28	1.64	1.96	2.33	2.58

Table B.3

df for denom.	1 - α	df for numerator											
		1	2	3	4	5	6	7	8	9	10	11	12
1	.75	5.83	7.50	8.20	8.58	8.82	8.98	9.10	9.19	9.26	9.32	9.36	9.41
	.90	39.9	49.5	53.6	55.8	57.2	58.2	58.9	59.4	59.9	60.2	60.5	60.7
	.95	161	200	216	225	230	234	237	239	241	242	243	244
	.99												
2	.75	2.57	3.00	3.15	3.23	3.28	3.31	3.34	3.35	3.37	3.38	3.39	3.39
	.90	8.53	9.00	9.16	9.24	9.29	9.33	9.35	9.37	9.38	9.39	9.40	9.41
	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4
3	.75	2.02	2.28	2.36	2.39	2.41	2.42	2.43	2.44	2.44	2.44	2.45	2.45
	.90	5.54	5.46	5.39	5.34	5.31	5.28	5.27	5.25	5.24	5.23	5.22	5.22
	.95	10.1	9.55	9.28	9.12	9.10	8.94	8.89	8.85	8.81	8.79	8.76	8.74
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1
4	.75	1.81	2.00	2.05	2.06	2.07	2.08	2.08	2.08	2.08	2.08	2.08	2.08
	.90	4.54	4.32	4.19	4.11	4.05	4.01	3.98	3.95	3.94	3.92	3.91	3.90
	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.4	14.4
5	.75	1.69	1.85	1.88	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89	1.89
	.90	4.06	3.78	3.62	3.52	3.45	3.40	3.37	3.34	3.32	3.30	3.28	3.27
	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.71	4.68
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89
6	.75	1.62	1.76	1.78	1.79	1.79	1.78	1.78	1.77	1.77	1.77	1.77	1.77
	.90	3.78	3.46	3.29	3.18	3.11	3.05	3.01	2.98	2.96	2.94	2.92	2.90
	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00
	.99	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	.75	1.57	1.70	1.72	1.72	1.71	1.71	1.70	1.70	1.69	1.69	1.69	1.68
	.90	3.59	3.26	3.07	2.96	2.88	2.83	2.78	2.75	2.72	2.70	2.68	2.67
	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47
8	.75	1.54	1.66	1.67	1.66	1.66	1.65	1.64	1.64	1.64	1.63	1.63	1.62
	.90	3.46	3.11	2.92	2.81	2.73	2.67	2.62	2.59	2.56	2.54	2.52	2.50
	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67
9	.75	1.51	1.62	1.63	1.63	1.62	1.61	1.60	1.60	1.59	1.59	1.58	1.58
	.90	3.36	3.01	2.81	2.69	2.61	2.55	2.51	2.47	2.44	2.42	2.40	2.38
	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11
10	.75	1.49	1.60	1.60	1.59	1.59	1.58	1.57	1.56	1.56	1.55	1.55	1.54
	.90	3.28	2.92	2.73	2.61	2.52	2.46	2.41	2.38	2.35	2.32	2.30	2.28
	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71
11	.75	1.47	1.58	1.58	1.57	1.56	1.55	1.54	1.53	1.53	1.52	1.52	1.51
	.90	3.23	2.86	2.66	2.54	2.45	2.39	2.34	2.30	2.27	2.25	2.23	2.21
	.95	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79
	.99	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40
12	.75	1.46	1.56	1.56	1.55	1.54	1.53	1.52	1.51	1.51	1.50	1.50	1.49
	.90	3.18	2.81	2.61	2.48	2.39	2.33	2.28	2.24	2.21	2.19	2.17	2.15
	.95	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69
	.99	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16

F Distribution*

df for numerator												1 - α	df for denom.
15	20	24	30	40	50	60	100	120	200	500	∞		
9.49	9.58	9.63	9.67	9.71	9.74	9.76	9.78	9.80	9.82	9.84	9.85	.75	1
61.2	61.7	62.0	62.3	62.5	62.7	62.8	63.0	63.1	63.2	63.3	63.3	.90	
246	248	249	250	251	252	252	253	253	254	254	254	.95	
3.41	3.43	3.43	3.44	3.45	3.45	3.46	3.47	3.47	3.48	3.48	3.48	.75	2
9.42	9.44	9.45	9.46	9.47	9.47	9.47	9.48	9.48	9.49	9.49	9.49	.90	
19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5	.95	
99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5	.99	3
2.46	2.46	2.46	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	2.47	.75	
5.20	5.18	5.18	5.17	5.16	5.15	5.15	5.14	5.14	5.14	5.14	5.13	.90	
8.70	8.66	8.64	8.62	8.59	8.58	8.57	8.55	8.55	8.54	8.53	8.53	.95	4
26.9	26.7	26.6	26.5	26.4	26.4	26.3	26.2	26.2	26.2	26.1	26.1	.99	
2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	2.08	.75	5
3.87	3.84	3.83	3.82	3.80	3.80	3.79	3.78	3.78	3.77	3.76	3.76	.90	
5.86	5.80	5.77	5.75	5.72	5.70	5.69	5.66	5.66	5.65	5.64	5.63	.95	
14.2	14.0	13.9	13.8	13.7	13.7	13.7	13.6	13.6	13.5	13.5	13.5	.99	6
1.89	1.88	1.88	1.88	1.88	1.88	1.87	1.87	1.87	1.87	1.87	1.87	.75	
3.24	3.21	3.19	3.17	3.16	3.15	3.14	3.13	3.12	3.12	3.11	3.10	.90	
4.62	4.56	4.53	4.50	4.46	4.44	4.43	4.41	4.40	4.39	4.37	4.36	.95	7
9.72	9.55	9.47	9.38	9.29	9.24	9.20	9.13	9.11	9.08	9.04	9.02	.99	
1.76	1.76	1.75	1.75	1.75	1.75	1.74	1.74	1.74	1.74	1.74	1.74	.75	
2.87	2.84	2.82	2.80	2.78	2.77	2.76	2.75	2.74	2.73	2.73	2.72	.90	8
3.94	3.87	3.84	3.81	3.77	3.75	3.74	3.71	3.70	3.69	3.68	3.67	.95	
7.56	7.40	7.31	7.23	7.14	7.09	7.06	6.99	6.97	6.93	6.90	6.88	.99	
1.68	1.67	1.67	1.66	1.66	1.66	1.65	1.65	1.65	1.65	1.65	1.65	.75	9
2.63	2.59	2.58	2.56	2.54	2.52	2.51	2.50	2.49	2.48	2.48	2.47	.90	
3.51	3.44	3.41	3.38	3.34	3.32	3.30	3.27	3.27	3.25	3.24	3.23	.95	
6.31	6.16	6.07	5.99	5.91	5.86	5.82	5.75	5.74	5.70	5.67	5.65	.99	10
1.62	1.61	1.60	1.60	1.59	1.59	1.59	1.58	1.58	1.58	1.58	1.58	.75	
2.46	2.42	2.40	2.38	2.36	2.35	2.34	2.32	2.32	2.31	2.30	2.29	.90	
3.22	3.15	3.12	3.08	3.04	3.02	3.01	2.97	2.97	2.95	2.94	2.93	.95	11
5.52	5.36	5.28	5.20	5.12	5.07	5.03	4.96	4.95	4.91	4.88	4.86	.99	
1.57	1.56	1.56	1.55	1.55	1.54	1.54	1.53	1.53	1.53	1.53	1.53	.75	
2.34	2.30	2.28	2.25	2.23	2.22	2.21	2.19	2.18	2.17	2.17	2.16	.90	12
3.01	2.94	2.90	2.86	2.83	2.80	2.79	2.76	2.75	2.73	2.72	2.71	.95	
4.96	4.81	4.73	4.65	4.57	4.52	4.48	4.42	4.40	4.36	4.33	4.31	.99	
1.53	1.52	1.52	1.51	1.51	1.50	1.50	1.49	1.49	1.49	1.48	1.48	.75	13
2.24	2.20	2.18	2.16	2.13	2.12	2.11	2.09	2.08	2.07	2.06	2.06	.90	
2.85	2.77	2.74	2.70	2.66	2.64	2.62	2.59	2.58	2.56	2.55	2.54	.95	
4.56	4.41	4.33	4.25	4.17	4.12	4.08	4.01	4.00	3.96	3.93	3.91	.99	14
1.50	1.49	1.49	1.48	1.47	1.47	1.47	1.46	1.46	1.46	1.45	1.45	.75	
2.17	2.12	2.10	2.08	2.05	2.04	2.03	2.00	2.00	1.99	1.98	1.97	.90	
2.72	2.65	2.61	2.57	2.53	2.51	2.49	2.46	2.45	2.43	2.42	2.40	.95	15
4.25	4.10	4.02	3.94	3.86	3.81	3.78	3.71	3.69	3.66	3.62	3.60	.99	
1.48	1.47	1.46	1.45	1.45	1.44	1.44	1.43	1.43	1.43	1.42	1.42	.75	
2.10	2.06	2.04	2.01	1.99	1.97	1.96	1.94	1.93	1.92	1.91	1.90	.90	16
2.62	2.54	2.51	2.47	2.43	2.40	2.38	2.35	2.34	2.32	2.31	2.30	.95	
4.01	3.86	3.78	3.70	3.62	3.57	3.54	3.47	3.45	3.41	3.38	3.36	.99	

Table B.3

df for denom.	$1 - \alpha$	df for numerator											
		1	2	3	4	5	6	7	8	9	10	11	12
13	.75	1.45	1.54	1.54	1.53	1.52	1.51	1.50	1.49	1.49	1.48	1.47	1.47
	.90	3.14	2.76	2.56	2.43	2.35	2.28	2.23	2.20	2.16	2.14	2.12	2.10
	.95	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60
	.99	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	.75	1.44	1.53	1.53	1.52	1.51	1.50	1.48	1.48	1.47	1.46	1.46	1.45
	.90	3.10	2.73	2.52	2.39	2.31	2.24	2.19	2.15	2.12	2.10	2.08	2.05
	.95	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53
	.99	8.86	6.51	5.56	5.04	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	.75	1.43	1.52	1.52	1.51	1.49	1.48	1.47	1.46	1.46	1.45	1.44	1.44
	.90	3.07	2.70	2.49	2.36	2.27	2.21	2.16	2.12	2.09	2.06	2.04	2.02
	.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	.75	1.42	1.51	1.51	1.50	1.48	1.48	1.47	1.46	1.45	1.45	1.44	1.44
	.90	3.05	2.67	2.46	2.33	2.24	2.18	2.13	2.09	2.06	2.03	2.01	1.99
	.95	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42
	.99	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55
17	.75	1.42	1.51	1.50	1.49	1.47	1.46	1.45	1.44	1.43	1.43	1.42	1.41
	.90	3.03	2.64	2.44	2.31	2.22	2.15	2.10	2.06	2.03	2.00	1.98	1.96
	.95	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38
	.99	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46
18	.75	1.41	1.50	1.49	1.48	1.46	1.45	1.44	1.43	1.42	1.42	1.41	1.40
	.90	3.01	2.62	2.42	2.29	2.20	2.13	2.08	2.04	2.00	1.98	1.96	1.93
	.95	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34
	.99	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37
19	.75	1.41	1.49	1.49	1.47	1.46	1.44	1.43	1.42	1.41	1.41	1.40	1.40
	.90	2.99	2.61	2.40	2.27	2.18	2.11	2.06	2.02	1.98	1.96	1.94	1.91
	.95	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31
	.99	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30
20	.75	1.40	1.49	1.48	1.46	1.45	1.44	1.42	1.42	1.41	1.40	1.39	1.39
	.90	2.97	2.59	2.38	2.25	2.16	2.09	2.04	2.00	1.96	1.94	1.92	1.89
	.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23
22	.75	1.40	1.48	1.47	1.45	1.44	1.42	1.41	1.40	1.39	1.39	1.38	1.37
	.90	2.95	2.56	2.35	2.22	2.13	2.06	2.01	1.97	1.93	1.90	1.88	1.86
	.95	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23
	.99	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12
24	.75	1.39	1.47	1.46	1.44	1.43	1.41	1.40	1.39	1.38	1.38	1.37	1.36
	.90	2.93	2.54	2.33	2.19	2.10	2.04	1.98	1.94	1.91	1.88	1.85	1.83
	.95	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.21	2.18
	.99	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03
26	.75	1.38	1.46	1.45	1.44	1.42	1.41	1.40	1.39	1.37	1.37	1.36	1.35
	.90	2.91	2.52	2.31	2.17	2.08	2.01	1.96	1.92	1.88	1.86	1.84	1.81
	.95	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18	2.15
	.99	7.72	5.53	4.64	4.14	3.82	3.59	3.42	3.29	3.18	3.09	3.02	2.96
28	.75	1.38	1.46	1.45	1.43	1.41	1.40	1.39	1.38	1.37	1.36	1.35	1.34
	.90	2.89	2.50	2.29	2.16	2.06	2.00	1.94	1.90	1.87	1.84	1.81	1.79
	.95	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15	2.12
	.99	7.64	5.45	4.57	4.07	3.75	3.53	3.36	3.23	3.12	3.03	2.96	2.90

F Distribution (Continued)*

df for numerator												1 - α	df for denom.
15	20	24	30	40	50	60	100	120	200	500	∞		
1.46	1.45	1.44	1.43	1.42	1.42	1.42	1.41	1.41	1.40	1.40	1.40	.75	13
2.05	2.01	1.98	1.96	1.93	1.92	1.90	1.88	1.88	1.86	1.85	1.85	.90	
2.53	2.46	2.42	2.38	2.34	2.31	2.30	2.26	2.25	2.23	2.22	2.21	.95	
3.82	3.66	3.59	3.51	3.43	3.38	3.34	3.27	3.25	3.22	3.19	3.17	.99	
1.44	1.43	1.42	1.41	1.41	1.40	1.40	1.39	1.39	1.39	1.38	1.38	.75	14
2.01	1.96	1.94	1.91	1.89	1.87	1.86	1.83	1.83	1.82	1.80	1.80	.90	
2.46	2.39	2.35	2.31	2.27	2.24	2.22	2.19	2.18	2.16	2.14	2.13	.95	
3.66	3.51	3.43	3.35	3.27	3.22	3.18	3.11	3.09	3.06	3.03	3.00	.99	
1.43	1.41	1.41	1.40	1.39	1.39	1.38	1.38	1.37	1.37	1.36	1.36	.75	15
1.97	1.92	1.90	1.87	1.85	1.83	1.82	1.79	1.79	1.77	1.76	1.76	.90	
2.40	2.33	2.29	2.25	2.20	2.18	2.16	2.12	2.11	2.10	2.08	2.07	.95	
3.52	3.37	3.29	3.21	3.13	3.08	3.05	2.98	2.96	2.92	2.89	2.87	.99	
1.41	1.40	1.39	1.38	1.37	1.37	1.36	1.36	1.35	1.35	1.34	1.34	.75	16
1.94	1.89	1.87	1.84	1.81	1.79	1.78	1.76	1.75	1.74	1.73	1.72	.90	
2.35	2.28	2.24	2.19	2.15	2.12	2.11	2.07	2.06	2.04	2.02	2.01	.95	
3.41	3.26	3.18	3.10	3.02	2.97	2.93	2.86	2.84	2.81	2.78	2.75	.99	
1.40	1.39	1.38	1.37	1.36	1.35	1.35	1.34	1.34	1.34	1.33	1.33	.75	17
1.91	1.86	1.84	1.81	1.78	1.76	1.75	1.73	1.72	1.71	1.69	1.69	.90	
2.31	2.23	2.19	2.15	2.10	2.08	2.06	2.02	2.01	1.99	1.97	1.96	.95	
3.31	3.16	3.08	3.00	2.92	2.87	2.83	2.76	2.75	2.71	2.68	2.65	.99	
1.39	1.38	1.37	1.36	1.35	1.34	1.34	1.33	1.33	1.32	1.32	1.32	.75	18
1.89	1.84	1.81	1.78	1.75	1.74	1.72	1.70	1.69	1.68	1.67	1.66	.90	
2.27	2.19	2.15	2.11	2.06	2.04	2.02	1.98	1.97	1.95	1.93	1.92	.95	
3.23	3.08	3.00	2.92	2.84	2.78	2.75	2.68	2.66	2.62	2.59	2.57	.99	
1.38	1.37	1.36	1.35	1.34	1.33	1.33	1.32	1.32	1.31	1.31	1.30	.75	19
1.86	1.81	1.79	1.76	1.73	1.71	1.70	1.67	1.67	1.65	1.64	1.63	.90	
2.23	2.16	2.11	2.07	2.03	2.00	1.98	1.94	1.93	1.91	1.89	1.88	.95	
3.15	3.00	2.92	2.84	2.76	2.71	2.67	2.60	2.58	2.55	2.51	2.49	.99	
1.37	1.36	1.35	1.34	1.33	1.33	1.32	1.31	1.31	1.30	1.30	1.29	.75	20
1.84	1.79	1.77	1.74	1.71	1.69	1.68	1.65	1.64	1.63	1.62	1.61	.90	
2.20	2.12	2.08	2.04	1.99	1.97	1.95	1.91	1.90	1.88	1.86	1.84	.95	
3.09	2.94	2.86	2.78	2.69	2.64	2.61	2.54	2.52	2.48	2.44	2.42	.99	
1.36	1.34	1.33	1.32	1.31	1.31	1.30	1.30	1.30	1.29	1.29	1.28	.75	22
1.81	1.76	1.73	1.70	1.67	1.65	1.64	1.61	1.60	1.59	1.58	1.57	.90	
2.15	2.07	2.03	1.98	1.94	1.91	1.89	1.85	1.84	1.82	1.80	1.78	.95	
2.98	2.83	2.75	2.67	2.58	2.53	2.50	2.42	2.40	2.36	2.33	2.31	.99	
1.35	1.33	1.32	1.31	1.30	1.29	1.29	1.28	1.28	1.27	1.27	1.26	.75	24
1.78	1.73	1.70	1.67	1.64	1.62	1.61	1.58	1.57	1.56	1.54	1.53	.90	
2.11	2.03	1.98	1.94	1.89	1.86	1.84	1.80	1.79	1.77	1.75	1.73	.95	
2.89	2.74	2.66	2.58	2.49	2.44	2.40	2.33	2.31	2.27	2.24	2.21	.99	
1.34	1.32	1.31	1.30	1.29	1.28	1.28	1.26	1.26	1.26	1.25	1.25	.75	26
1.76	1.71	1.68	1.65	1.61	1.59	1.58	1.55	1.54	1.53	1.51	1.50	.90	
2.07	1.99	1.95	1.90	1.85	1.82	1.80	1.76	1.75	1.73	1.71	1.69	.95	
2.81	2.66	2.58	2.50	2.42	2.36	2.33	2.25	2.23	2.19	2.16	2.13	.99	
1.33	1.31	1.30	1.29	1.28	1.27	1.27	1.26	1.25	1.25	1.24	1.24	.75	28
1.74	1.69	1.66	1.63	1.59	1.57	1.56	1.53	1.52	1.50	1.49	1.48	.90	
2.04	1.96	1.91	1.87	1.82	1.79	1.77	1.73	1.71	1.69	1.67	1.65	.95	
2.75	2.60	2.52	2.44	2.35	2.30	2.26	2.19	2.17	2.13	2.09	2.06	.99	

Table B.3

df for denom.	$1 - \alpha$	df for numerator											
		1	2	3	4	5	6	7	8	9	10	11	12
30	.75	1.38	1.45	1.44	1.42	1.41	1.39	1.38	1.37	1.36	1.35	1.35	1.34
	.90	2.88	2.49	2.28	2.14	2.05	1.98	1.93	1.88	1.85	1.82	1.79	1.77
	.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84
40	.75	1.36	1.44	1.42	1.40	1.39	1.37	1.36	1.35	1.34	1.33	1.32	1.31
	.90	2.84	2.44	2.23	2.09	2.00	1.93	1.87	1.83	1.79	1.76	1.73	1.71
	.95	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00
	.99	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66
60	.75	1.35	1.42	1.41	1.38	1.37	1.35	1.33	1.32	1.31	1.30	1.29	1.29
	.90	2.79	2.39	2.18	2.04	1.95	1.87	1.82	1.77	1.74	1.71	1.68	1.66
	.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50
120	.75	1.34	1.40	1.39	1.37	1.35	1.33	1.31	1.30	1.29	1.28	1.27	1.26
	.90	2.75	2.35	2.13	1.99	1.90	1.82	1.77	1.72	1.68	1.65	1.62	1.60
	.95	3.92	3.07	2.68	2.45	2.29	2.17	2.09	2.02	1.96	1.91	1.87	1.83
	.99	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.40	2.34
200	.75	1.33	1.39	1.38	1.36	1.34	1.32	1.31	1.29	1.28	1.27	1.26	1.25
	.90	2.73	2.33	2.11	1.97	1.88	1.80	1.75	1.70	1.66	1.63	1.60	1.57
	.95	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80
	.99	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27
∞	.75	1.32	1.39	1.37	1.35	1.33	1.31	1.29	1.28	1.27	1.25	1.24	1.24
	.90	2.71	2.30	2.08	1.94	1.85	1.77	1.72	1.67	1.63	1.60	1.57	1.55
	.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18

* This table is abridged from Table 18 in *Biometrika Tables for Statisticians*, vol. 1. Reproduced with the kind permission of E. S. Pearson and the trustees of *Biometrika*.

F Distribution (Continued)*

df for numerator												1 - α	df for denom.
15	20	24	30	40	50	60	100	120	200	500	∞		
1.32	1.30	1.29	1.28	1.27	1.26	1.26	1.25	1.24	1.24	1.23	1.23	.75	30
1.72	1.67	1.64	1.61	1.57	1.55	1.54	1.51	1.50	1.48	1.47	1.46	.90	
2.01	1.93	1.89	1.84	1.79	1.76	1.74	1.70	1.68	1.66	1.64	1.62	.95	
2.70	2.55	2.47	2.39	2.30	2.25	2.21	2.13	2.11	2.07	2.03	2.01	.99	
1.30	1.28	1.26	1.25	1.24	1.23	1.22	1.21	1.21	1.20	1.19	1.19	.75	40
1.66	1.61	1.57	1.54	1.51	1.48	1.47	1.43	1.42	1.41	1.39	1.38	.90	
1.92	1.84	1.79	1.74	1.69	1.66	1.64	1.59	1.58	1.55	1.53	1.51	.95	
2.52	2.37	2.29	2.20	2.11	2.06	2.02	1.94	1.92	1.87	1.83	1.80	.99	
1.27	1.25	1.24	1.22	1.21	1.20	1.19	1.17	1.17	1.16	1.15	1.15	.75	60
1.60	1.54	1.51	1.48	1.44	1.41	1.40	1.36	1.35	1.33	1.31	1.29	.90	
1.84	1.75	1.70	1.65	1.59	1.56	1.53	1.48	1.47	1.44	1.41	1.39	.95	
2.35	2.20	2.12	2.03	1.94	1.88	1.84	1.75	1.73	1.68	1.63	1.60	.99	
1.24	1.22	1.21	1.19	1.18	1.17	1.16	1.14	1.13	1.12	1.11	1.10	.75	120
1.55	1.48	1.45	1.41	1.37	1.34	1.32	1.27	1.26	1.24	1.21	1.19	.90	
1.75	1.66	1.61	1.55	1.50	1.46	1.43	1.37	1.35	1.32	1.28	1.25	.95	
2.19	2.03	1.95	1.86	1.76	1.70	1.66	1.56	1.53	1.48	1.42	1.38	.99	
1.23	1.21	1.20	1.18	1.16	1.14	1.12	1.11	1.10	1.09	1.08	1.06	.75	200
1.52	1.46	1.42	1.38	1.34	1.31	1.28	1.24	1.22	1.20	1.17	1.14	.90	
1.72	1.62	1.57	1.52	1.46	1.41	1.39	1.32	1.29	1.26	1.22	1.19	.95	
2.13	1.97	1.89	1.79	1.69	1.63	1.58	1.48	1.44	1.39	1.33	1.28	.99	
1.22	1.19	1.18	1.16	1.14	1.13	1.12	1.09	1.08	1.07	1.04	1.00	.75	∞
1.49	1.42	1.38	1.34	1.30	1.26	1.24	1.18	1.17	1.13	1.08	1.00	.90	
1.67	1.57	1.52	1.46	1.39	1.35	1.32	1.24	1.22	1.17	1.11	1.00	.95	
2.04	1.88	1.79	1.70	1.59	1.52	1.47	1.36	1.32	1.25	1.15	1.00	.99	

(2d ed.) New York: Cambridge, 1958. Edited by E. S. Pearson and H. O. Hartley

Table B.4 Distribution of the Studentized Range Statistic*

df for s_x^2		$1 - \alpha$	r = number of steps between ordered means														
			2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	.95 .99	18.0 90.0	27.0 135	32.8 164	37.1 186	40.4 202	43.1 216	45.4 227	47.4 237	49.1 246	50.6 253	52.0 260	53.2 266	54.3 272	55.4 277		
		6.09 14.0	8.3 19.0	9.8 22.3	10.9 24.7	11.7 26.6	12.4 28.2	13.0 29.5	13.5 30.7	14.0 31.7	14.4 32.6	14.7 33.4	15.1 34.1	15.4 34.8	15.7 35.4		
3	.95 .99	4.50 8.26	5.91 10.6	6.82 12.2	7.50 13.3	8.04 14.2	8.48 15.0	8.85 15.6	9.18 16.2	9.46 16.7	9.72 17.1	9.95 17.5	10.2 17.9	10.4 18.2	10.5 18.5		
		3.93 6.51	5.04 8.12	5.76 9.17	6.29 9.96	6.71 10.6	7.05 11.1	7.35 11.5	7.60 11.9	7.83 12.3	8.03 12.6	8.21 12.8	8.37 13.1	8.52 13.3	8.66 13.5		
5	.95 .99	3.64 5.70	4.60 6.97	5.22 7.80	5.67 8.42	6.03 8.91	6.33 9.32	6.58 9.67	6.80 9.97	6.99 10.2	7.17 10.5	7.32 10.7	7.47 10.9	7.60 11.1	7.72 11.2		
		3.46 5.24	4.34 6.33	4.90 7.03	5.31 7.56	5.63 7.97	5.89 8.32	6.12 8.61	6.32 8.87	6.49 9.10	6.65 9.30	6.79 9.49	6.92 9.65	7.03 9.81	7.14 9.95		
7	.95 .99	3.34 4.95	4.16 5.92	4.69 6.54	5.06 7.01	5.36 7.37	5.61 7.68	5.82 7.94	6.00 8.17	6.16 8.37	6.30 8.55	6.43 8.71	6.55 8.86	6.66 9.00	6.76 9.12		
		3.26 4.74	4.04 5.63	4.53 6.20	4.89 6.63	5.17 6.96	5.40 7.24	5.60 7.47	5.77 7.68	5.92 7.87	6.05 8.03	6.18 8.18	6.29 8.31	6.39 8.44	6.48 8.55		
9	.95 .99	3.20 4.60	3.95 5.43	4.42 5.96	4.76 6.35	5.02 6.66	5.24 6.91	5.43 7.13	5.60 7.32	5.74 7.49	5.87 7.65	5.98 7.78	6.09 7.91	6.19 8.03	6.28 8.13		
		3.15 4.48	3.88 5.27	4.33 5.77	4.65 6.14	4.91 6.43	5.12 6.67	5.30 6.87	5.46 7.05	5.60 7.21	5.72 7.36	5.83 7.48	5.93 7.60	6.03 7.71	6.11 7.81		
11	.95 .99	3.11 4.39	3.82 5.14	4.26 5.62	4.57 5.97	4.82 6.25	5.03 6.48	5.20 6.67	5.35 6.84	5.49 6.99	5.61 7.13	5.71 7.26	5.81 7.36	5.90 7.46	5.99 7.56		
		3.08 4.32	3.77 5.04	4.20 5.50	4.51 5.84	4.75 6.10	4.95 6.32	5.12 6.51	5.27 6.67	5.40 6.81	5.51 6.94	5.62 7.06	5.71 7.17	5.80 7.26	5.88 7.36		

13	.95 .99	3.06 4.26	3.73 4.96	4.15 5.40	4.45 5.73	4.69 5.98	4.88 6.19	5.05 6.37	5.19 6.53	5.32 6.67	5.43 6.79	5.53 6.90	5.63 7.01	5.71 7.10	5.79 7.19
14	.95 .99	3.03 4.21	3.70 4.89	4.11 5.32	4.41 5.63	4.64 5.88	4.83 6.08	4.99 6.26	5.13 6.41	5.25 6.54	5.36 6.66	5.46 6.77	5.55 6.87	5.64 6.96	5.72 7.05
16	.95 .99	3.00 4.13	3.65 4.78	4.05 5.19	4.33 5.49	4.56 5.72	4.74 5.92	4.90 6.08	5.03 6.22	5.15 6.35	5.26 6.46	5.35 6.56	5.44 6.66	5.52 6.74	5.59 6.82
18	.95 .99	2.97 4.07	3.61 4.70	4.00 5.09	4.28 5.38	4.49 5.60	4.67 5.79	4.82 5.94	4.96 6.08	5.07 6.20	5.17 6.31	5.27 6.41	5.35 6.50	5.43 6.58	5.50 6.65
20	.95 .99	2.95 4.02	3.58 4.64	3.96 5.02	4.23 5.29	4.45 5.51	4.62 5.69	4.77 5.84	4.90 5.97	5.01 6.09	5.11 6.19	5.20 6.29	5.28 6.37	5.36 6.45	5.43 6.52
24	.95 .99	2.92 3.96	3.53 4.54	3.90 4.91	4.17 5.17	4.37 5.37	4.54 5.54	4.68 5.69	4.81 5.81	4.92 5.92	5.01 6.02	5.10 6.11	5.18 6.19	5.25 6.26	5.32 6.33
30	.95 .99	2.89 3.89	3.49 4.45	3.84 4.80	4.10 5.05	4.30 5.24	4.46 5.40	4.60 5.54	4.72 5.56	4.83 5.76	4.92 5.85	5.00 5.93	5.08 6.01	5.15 6.08	5.21 6.14
40	.95 .99	2.86 3.82	3.44 4.37	3.79 4.70	4.04 4.93	4.23 5.11	4.39 5.27	4.52 5.39	4.63 5.50	4.74 5.60	4.82 5.69	4.91 5.77	4.98 5.84	5.05 5.90	5.11 5.96
60	.95 .99	2.83 3.76	3.40 4.28	3.74 4.60	3.98 4.82	4.16 4.99	4.31 5.13	4.44 5.25	4.55 5.36	4.65 5.45	4.73 5.53	4.81 5.60	4.88 5.67	4.94 5.73	5.00 5.79
120	.95 .99	2.80 3.70	3.36 4.20	3.69 4.50	3.92 4.71	4.10 4.87	4.24 5.01	4.36 5.12	4.48 5.21	4.56 5.30	4.64 5.38	4.72 5.44	4.78 5.51	4.84 5.56	4.90 5.61
∞	.95 .99	2.77 3.64	3.31 4.12	3.63 4.40	3.86 4.60	4.03 4.76	4.17 4.88	4.29 4.99	4.39 5.08	4.47 5.16	4.55 5.23	4.62 5.29	4.68 5.35	4.74 5.40	4.80 5.45

* This table is abridged from Table 11.2 in *The Probability Integrals of the Range and of the Studentized Range*, prepared by H. Leon Harter, Donald S. Clemm, and Eugene H. Guthrie. These tables are published in WADC tech. Rep. 58-484, vol. 2, 1959, Wright Air Development Center, and are reproduced with the kind permission of the authors.

Table B.5 Arcsin Transformation ($\phi = 2 \arcsin \sqrt{X}$)

X	ϕ	X	ϕ	X	ϕ	X	ϕ	X	ϕ
.001	.0633	.041	.4078	.36	1.2870	.76	2.1177	.971	2.7993
.002	.0895	.042	.4128	.37	1.3078	.77	2.1412	.972	2.8053
.003	.1096	.043	.4178	.38	1.3284	.78	2.1652	.973	2.8115
.004	.1266	.044	.4227	.39	1.3490	.79	2.1895	.974	2.8177
.005	.1415	.045	.4275	.40	1.3694	.80	2.2143	.975	2.8240
.006	.1551	.046	.4323	.41	1.3898	.81	2.2395	.976	2.8305
.007	.1675	.047	.4371	.42	1.4101	.82	2.2653	.977	2.8371
.008	.1791	.048	.4418	.43	1.4303	.83	2.2916	.978	2.8438
.009	.1900	.049	.4464	.44	1.4505	.84	2.3186	.979	2.8507
.010	.2003	.050	.4510	.45	1.4706	.85	2.3462	.980	2.8578
.011	.2101	.06	.4949	.46	1.4907	.86	2.3746	.981	2.8650
.012	.2195	.07	.5355	.47	1.5108	.87	2.4039	.982	2.8725
.013	.2285	.08	.5735	.48	1.5308	.88	2.4341	.983	2.8801
.014	.2372	.09	.6094	.49	1.5508	.89	2.4655	.984	2.8879
.015	.2456	.10	.6435	.50	1.5708	.90	2.4981	.985	2.8960
.016	.2537	.11	.6761	.51	1.5908	.91	2.5322	.986	2.9044
.017	.2615	.12	.7075	.52	1.6108	.92	2.5681	.987	2.9131
.018	.2691	.13	.7377	.53	1.6308	.93	2.6062	.988	2.9221
.019	.2766	.14	.7670	.54	1.6509	.94	2.6467	.989	2.9315
.020	.2838	.15	.7954	.55	1.6710	.95	2.6906	.990	2.9413
.021	.2909	.16	.8230	.56	1.6911	.951	2.6952	.991	2.9516
.022	.2978	.17	.8500	.57	1.7113	.952	2.6998	.992	2.9625
.023	.3045	.18	.8763	.58	1.7315	.953	2.7045	.993	2.9741
.024	.3111	.19	.9021	.59	1.7518	.954	2.7093	.994	2.9865
.025	.3176	.20	.9273	.60	1.7722	.955	2.7141	.995	3.0001
.026	.3239	.21	.9521	.61	1.7926	.956	2.7189	.996	3.0150
.027	.3301	.22	.9764	.62	1.8132	.957	2.7238	.997	3.0320
.028	.3363	.23	1.0004	.63	1.8338	.958	2.7288	.998	3.0521
.029	.3423	.24	1.0239	.64	1.8546	.959	2.7338	.999	3.0783
.030	.3482	.25	1.0472	.65	1.8755	.960	2.7389		
.031	.3540	.26	1.0701	.66	1.8965	.961	2.7440		
.032	.3597	.27	1.0928	.67	1.9177	.962	2.7492		
.033	.3654	.28	1.1152	.68	1.9391	.963	2.7545		
.034	.3709	.29	1.1374	.69	1.9606	.964	2.7598		
.035	.3764	.30	1.1593	.70	1.9823	.965	2.7652		
.036	.3818	.31	1.1810	.71	2.0042	.966	2.7707		
.037	.3871	.32	1.2025	.72	2.0264	.967	2.7762		
.038	.3924	.33	1.2239	.73	2.0488	.968	2.7819		
.039	.3976	.34	1.2451	.74	2.0715	.969	2.7876		
.040	.4027	.35	1.2661	.75	2.0944	.970	2.7934		

Table B.6 Distribution of t Statistic in Comparing Treatment Means with a Control

df for MS _{error}	$1 - \alpha$	$k = \text{number of means (including control)}$								
		2	3	4	5	6	7	8	9	10
5	.95	2.02	2.44	2.68	2.85	2.98	3.08	3.16	3.24	3.30
	.975	2.57	3.03	3.39	3.66	3.88	4.06	4.22	4.36	4.49
	.99	3.37	3.90	4.21	4.43	4.60	4.73	4.85	4.94	5.03
	.995	4.03	4.63	5.09	5.44	5.73	5.97	6.18	6.36	6.53
6	.95	1.94	2.34	2.56	2.71	2.83	2.92	3.00	3.07	3.12
	.975	2.45	2.86	3.18	3.41	3.60	3.75	3.88	4.00	4.11
	.99	3.14	3.61	3.88	4.07	4.21	4.33	4.43	4.51	4.59
	.995	3.71	4.22	4.60	4.88	5.11	5.30	5.47	5.61	5.74
7	.95	1.89	2.27	2.48	2.62	2.73	2.82	2.89	2.95	3.01
	.975	2.36	2.75	3.04	3.24	3.41	3.54	3.66	3.76	3.86
	.99	3.00	3.42	3.66	3.83	3.96	4.07	4.15	4.23	4.30
	.995	3.50	3.95	4.28	4.52	4.71	4.87	5.01	5.13	5.24
8	.95	1.86	2.22	2.42	2.55	2.66	2.74	2.81	2.87	2.92
	.975	2.31	2.67	2.94	3.13	3.28	3.40	3.51	3.60	3.68
	.99	2.90	3.29	3.51	3.67	3.79	3.88	3.96	4.03	4.09
	.995	3.36	3.77	4.06	4.27	4.44	4.58	4.70	4.81	4.90
9	.95	1.83	2.18	2.37	2.50	2.60	2.68	2.75	2.81	2.86
	.975	2.26	2.61	2.86	3.04	3.18	3.29	3.39	3.48	3.55
	.99	2.82	3.19	3.40	3.55	3.66	3.75	3.82	3.89	3.94
	.995	3.25	3.63	3.90	4.09	4.24	4.37	4.48	4.57	4.65
10	.95	1.81	2.15	2.34	2.47	2.56	2.64	2.70	2.76	2.81
	.975	2.23	2.57	2.81	2.97	3.11	3.21	3.31	3.39	3.46
	.99	2.76	3.11	3.31	3.45	3.56	3.64	3.71	3.78	3.83
	.995	3.17	3.53	3.78	3.95	4.10	4.21	4.31	4.40	4.47
11	.95	1.80	2.13	2.31	2.44	2.53	2.60	2.67	2.72	2.77
	.975	2.20	2.53	2.76	2.92	3.05	3.15	3.24	3.31	3.38
	.99	2.72	3.06	3.25	3.38	3.48	3.56	3.63	3.69	3.74
	.995	3.11	3.45	3.68	3.85	3.98	4.09	4.18	4.26	4.33
12	.95	1.78	2.11	2.29	2.41	2.50	2.58	2.64	2.69	2.74
	.975	2.18	2.50	2.72	2.88	3.00	3.10	3.18	3.25	3.32
	.99	2.68	3.01	3.19	3.32	3.42	3.50	3.56	3.62	3.67
	.995	3.05	3.39	3.61	3.76	3.89	3.99	4.08	4.15	4.22
13	.95	1.77	2.09	2.27	2.39	2.48	2.55	2.61	2.66	2.71
	.975	2.16	2.48	2.69	2.84	2.96	3.06	3.14	3.21	3.27
	.99	2.65	2.97	3.15	3.27	3.37	3.44	3.51	3.56	3.61
	.995	3.01	3.33	3.54	3.69	3.81	3.91	3.99	4.06	4.13
14	.95	1.76	2.08	2.25	2.37	2.46	2.53	2.59	2.64	2.69
	.975	2.14	2.46	2.67	2.81	2.93	3.02	3.10	3.17	3.23
	.99	2.62	2.94	3.11	3.23	3.32	3.40	3.46	3.51	3.56
	.995	2.98	3.29	3.49	3.64	3.75	3.84	3.92	3.99	4.05

Table B.6 Distribution of t Statistic when Comparing Treatment Means With a Control (*Continued*)*

df for MS _{Error}	1 - α	k = number of means (including control)								
		2	3	4	5	6	7	8	9	10
16	.95	1.75	2.06	2.23	2.34	2.43	2.50	2.56	2.61	2.65
	.975	2.12	2.42	2.63	2.77	2.88	2.96	3.04	3.10	3.16
	.99	2.58	2.88	3.05	3.17	3.26	3.33	3.39	3.44	3.48
	.995	2.92	3.22	3.41	3.55	3.65	3.74	3.82	3.88	3.93
18	.95	1.73	2.04	2.21	2.32	2.41	2.48	2.53	2.58	2.62
	.975	2.10	2.40	2.59	2.73	2.84	2.92	2.99	3.05	3.11
	.99	2.55	2.84	3.01	3.12	3.21	3.27	3.33	3.38	3.42
	.995	2.88	3.17	3.35	3.48	3.58	3.67	3.74	3.80	3.85
20	.95	1.72	2.03	2.19	2.30	2.39	2.46	2.51	2.56	2.60
	.975	2.09	2.38	2.57	2.70	2.81	2.89	2.96	3.02	3.07
	.99	2.53	2.81	2.97	3.08	3.17	3.23	3.29	3.34	3.38
	.995	2.85	3.13	3.31	3.43	3.53	3.61	3.67	3.73	3.78
24	.95	1.71	2.01	2.17	2.28	2.36	2.43	2.48	2.53	2.57
	.975	2.06	2.35	2.53	2.66	2.76	2.84	2.91	2.96	3.01
	.99	2.49	2.77	2.92	3.03	3.11	3.17	3.22	3.27	3.31
	.995	2.80	3.07	3.24	3.36	3.45	3.52	3.58	3.64	3.69
30	.95	1.70	1.99	2.15	2.25	2.33	2.40	2.45	2.50	2.54
	.975	2.04	2.32	2.50	2.62	2.72	2.79	2.86	2.91	2.96
	.99	2.46	2.72	2.87	2.97	3.05	3.11	3.16	3.21	3.24
	.995	2.75	3.01	3.17	3.28	3.37	3.44	3.50	3.55	3.59
40	.95	1.68	1.97	2.13	2.23	2.31	2.37	2.42	2.47	2.51
	.975	2.02	2.29	2.47	2.58	2.67	2.75	2.81	2.86	2.90
	.99	2.42	2.68	2.82	2.92	2.99	3.05	3.10	3.14	3.18
	.995	2.70	2.95	3.10	3.21	3.29	3.36	3.41	3.46	3.50
60	.95	1.67	1.95	2.10	2.21	2.28	2.35	2.39	2.44	2.48
	.975	2.00	2.27	2.43	2.55	2.63	2.70	2.76	2.81	2.85
	.99	2.39	2.64	2.78	2.87	2.94	3.00	3.04	3.08	3.12
	.995	2.66	2.90	3.04	3.14	3.22	3.28	3.33	3.38	3.42
120	.95	1.66	1.93	2.08	2.18	2.26	2.32	2.37	2.41	2.45
	.975	1.98	2.24	2.40	2.51	2.59	2.66	2.71	2.76	2.80
	.99	2.36	2.60	2.73	2.82	2.89	2.94	2.99	3.03	3.06
	.995	2.62	2.84	2.98	3.08	3.15	3.21	3.25	3.30	3.33
∞	.95	1.64	1.92	2.06	2.16	2.23	2.29	2.34	2.38	2.42
	.975	1.96	2.21	2.37	2.47	2.55	2.62	2.67	2.71	2.75
	.99	2.33	2.56	2.68	2.77	2.84	2.89	2.93	2.97	3.00
	.995	2.58	2.79	2.92	3.01	3.08	3.14	3.18	3.22	3.25

* This table is reproduced from: A multiple comparison procedure for comparing several treatments with a control. *J. Am. statist. Ass.*, 1955, 50, 1096-1121, with the permission of the author, C. W. Dunnett, and the editor.

Table B.7 Distribution of F_{\max} Statistic*

df for s_X^2	$1 - \alpha$	$k = \text{number of variances}$								
		2	3	4	5	6	7	8	9	10
4	.95	9.60	15.5	20.6	25.2	29.5	33.6	37.5	41.4	44.6
	.99	23.2	37.	49.	59.	69.	79.	89.	97.	106.
5	.95	7.15	10.8	13.7	16.3	18.7	20.8	22.9	24.7	26.5
	.99	14.9	22.	28.	33.	38.	42.	46.	50.	54.
6	.95	5.82	8.38	10.4	12.1	13.7	15.0	16.3	17.5	18.6
	.99	11.1	15.5	19.1	22.	25.	27.	30.	32.	34.
7	.95	4.99	6.94	8.44	9.70	10.8	11.8	12.7	13.5	14.3
	.99	8.89	12.1	14.5	16.5	18.4	20.	22.	23.	24.
8	.95	4.43	6.00	7.18	8.12	9.03	9.78	10.5	11.1	11.7
	.99	7.50	9.9	11.7	13.2	14.5	15.8	16.9	17.9	18.9
9	.95	4.03	5.34	6.31	7.11	7.80	8.41	8.95	9.45	9.91
	.99	6.54	8.5	9.9	11.1	12.1	13.1	13.9	14.7	15.3
10	.95	3.72	4.85	5.67	6.34	6.92	7.42	7.87	8.28	8.66
	.99	5.85	7.4	8.6	9.6	10.4	11.1	11.8	12.4	12.9
12	.95	3.28	4.16	4.79	5.30	5.72	6.09	6.42	6.72	7.00
	.99	4.91	6.1	6.9	7.6	8.2	8.7	9.1	9.5	9.9
15	.95	2.86	3.54	4.01	4.37	4.68	4.95	5.19	5.40	5.59
	.99	4.07	4.9	5.5	6.0	6.4	6.7	7.1	7.3	7.5
20	.95	2.46	2.95	3.29	3.54	3.76	3.94	4.10	4.24	4.37
	.99	3.32	3.8	4.3	4.6	4.9	5.1	5.3	5.5	5.6
30	.95	2.07	2.40	2.61	2.78	2.91	3.02	3.12	3.21	3.29
	.99	2.63	3.0	3.3	3.4	3.6	3.7	3.8	3.9	4.0
60	.95	1.67	1.85	1.96	2.04	2.11	2.17	2.22	2.26	2.30
	.99	1.96	2.2	2.3	2.4	2.4	2.5	2.5	2.6	2.6
∞	.95	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	.99	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

* This table is abridged from Table 31 in *Biometrika Tables for Statisticians*, vol. 1. (2d ed.) New York: Cambridge, 1958. Edited by E. S. Pearson and H. O. Hartley. Reproduced with the kind permission of E. S. Pearson and the trustees of *Biometrika*.

Table B.8 Critical Values for Cochran's Test for Homogeneity of Variance*

$$C = (\text{largest } s^2) / (\sum s_j^2)$$

df for s_j^2	$1 - \alpha$	$k = \text{number of variances}$											
		2	3	4	5	6	7	8	9	10	15	20	
1	.95	.9985	.9669	.9065	.8412	.7808	.7271	.6798	.6385	.6020	.4709	.3894	
	.99	.9999	.9933	.9676	.9279	.8828	.8376	.7945	.7544	.7175	.5747	.4799	
2	.95	.9750	.8709	.7679	.6838	.6161	.5612	.5157	.4775	.4450	.3346	.2705	
	.99	.9950	.9423	.8643	.7885	.7218	.6644	.6152	.5727	.5358	.4069	.3297	
3	.95	.9392	.7977	.6841	.5981	.5321	.4800	.4377	.4027	.3733	.2758	.2205	
	.99	.9794	.8831	.7814	.6957	.6258	.5685	.5209	.4810	.4469	.3317	.2654	
4	.95	.9057	.7457	.6287	.5441	.4803	.4307	.3910	.3584	.3311	.2419	.1921	
	.99	.9586	.8335	.7212	.6329	.5635	.5080	.4627	.4251	.3934	.2882	.2288	
5	.95	.8772	.7071	.5895	.5065	.4447	.3974	.3595	.3286	.3029	.2195	.1735	
	.99	.9373	.7933	.6761	.5875	.5195	.4659	.4226	.3870	.3572	.2593	.2048	
6	.95	.8534	.6771	.5598	.4783	.4184	.3726	.3362	.3067	.2823	.2034	.1602	
	.99	.9172	.7606	.6410	.5531	.4866	.4347	.3932	.3592	.3308	.2386	.1877	
7	.95	.8332	.6530	.5365	.4564	.3980	.3535	.3185	.2901	.2666	.1911	.1501	
	.99	.8988	.7335	.6129	.5259	.4608	.4105	.3704	.3378	.3106	.2228	.1748	
8	.95	.8159	.6333	.5175	.4387	.3817	.3384	.3043	.2768	.2541	.1815	.1422	
	.99	.8823	.7107	.5897	.5037	.4401	.3911	.3522	.3207	.2945	.2104	.1646	
9	.95	.8010	.6167	.5017	.4241	.3682	.3259	.2926	.2659	.2439	.1736	.1357	
	.99	.8674	.6912	.5702	.4854	.4229	.3751	.3373	.3067	.2813	.2002	.1567	
16	.95	.7341	.5466	.4366	.3645	.3135	.2756	.2462	.2226	.2032	.1429	.1108	
	.99	.7949	.6059	.4884	.4094	.3529	.3105	.2779	.2514	.2297	.1612	.1248	
36	.95	.6602	.4748	.3720	.3066	.2612	.2278	.2022	.1820	.1655	.1144	.0879	
	.99	.7067	.5153	.4057	.3351	.2858	.2494	.2214	.1992	.1811	.1251	.0960	
144	.95	.5813	.4031	.3093	.2513	.2119	.1833	.1616	.1446	.1308	.0889	.0675	
	.99	.6062	.4230	.3251	.2644	.2229	.1929	.1700	.1521	.1376	.0934	.0709	

* Reproduced with permission from C. Eisenhart, M. W. Hastay, and W. A. Wallis, *Techniques of Statistical Analysis*, chap. 15. New York: McGraw-Hill, 1947.

Table B.9 Chi-square Distribution*

df	Percentile point						
	50	75	90	95	97.5	99	99.5
1	.46	1.3	2.7	3.8	5.0	6.6	7.9
2	1.4	2.8	4.6	6.0	7.4	9.2	10.6
3	2.4	4.1	6.3	7.8	9.4	11.3	12.8
4	3.4	5.4	7.8	9.5	11.1	13.3	14.9
5	4.4	6.6	9.2	11.1	12.8	15.1	16.7
6	5.4	7.8	10.6	12.6	14.4	16.8	18.5
7	6.4	9.0	12.0	14.1	16.0	18.5	20.3
8	7.3	10.2	13.4	15.5	17.5	20.1	22.0
9	8.3	11.4	14.7	16.9	19.0	21.7	23.6
10	9.3	12.5	16.0	18.3	20.5	23.2	25.2
11	10.3	13.7	17.3	19.7	21.9	24.7	26.8
12	11.3	14.8	18.5	21.0	23.3	26.2	28.3
13	12.3	16.0	19.8	22.4	24.7	27.7	29.8
14	13.3	17.1	21.1	23.7	26.1	29.1	31.3
15	14.3	18.2	22.3	25.0	27.5	30.6	32.8
16	15.3	19.4	23.5	26.3	28.8	32.0	34.3
17	16.3	20.5	24.8	27.6	30.2	33.4	35.7
18	17.3	21.6	26.0	28.9	31.5	34.8	37.2
19	18.3	22.7	27.2	30.1	32.9	36.2	38.6
20	19.3	23.8	28.4	31.4	34.2	37.6	40.0
21	20.3	24.9	29.6	32.7	35.5	38.9	41.4
22	21.3	26.0	30.8	33.9	36.8	40.3	42.8
23	22.3	27.1	32.0	35.2	38.1	41.6	44.2
24	23.3	28.2	33.2	36.4	39.4	43.0	45.6
25	24.3	29.3	34.4	37.7	40.6	44.3	46.9
26	25.3	30.4	35.6	38.9	41.9	45.6	48.3
27	26.3	31.5	36.7	40.1	43.2	47.0	49.6
28	27.3	32.6	37.9	41.3	44.5	48.3	51.0
29	28.3	33.7	39.1	42.6	45.7	49.6	52.3
30	29.3	34.8	40.3	43.8	47.0	50.9	53.7
40	39.3	45.6	51.8	55.8	59.3	63.7	66.8
60	59.3	67.0	74.4	79.1	83.3	88.4	92.0
100	99.3	109.1	118.5	124.3	129.6	135.8	140.2

For $df > 30$, $\chi^2_{1-\alpha} \doteq [\sqrt{2(df) - 1} + z_{1-\alpha}]^2/2$.

For example, when $df = 60$,

$$\chi^2_{.95} = [\sqrt{2(60) - 1} + 1.645]^2/2 = 79.$$

* This table is abridged from Table 8 in *Biometrika Tables for Statisticians*, vol. 1. (2d ed.) New York: Cambridge, 1958. Edited by E. S. Pearson and H. O. Hartley. Reproduced with the kind permission of E. S. Pearson and the trustees of *Biometrika*.

Table B.10 Coefficients of Orthogonal Polynomials

k	Polynomial	$X=1$	2	3	4	5	6	7	8	9	10	$\sum \xi^{k/2}$	λ
3	Linear	-1	0	1								2	1
	Quadratic	1	-2	1								6	3
4	Linear	-3	-1	1	3							20	2
	Quadratic	1	-1	-1	1							4	1
	Cubic	-1	3	-3	1							20	$10/3$
5	Linear	-2	-1	0	1	2						10	1
	Quadratic	2	-1	-2	-1	2						14	1
	Cubic	-1	2	0	-2	1						10	$5/6$
	Quartic	1	-4	6	-4	1						70	$35/12$
6	Linear	-5	-3	-1	1	3	5					70	2
	Quadratic	5	-1	-4	-4	-1	5					84	$3/2$
	Cubic	-5	7	4	-4	-7	5					180	$5/3$
	Quartic	1	-3	2	2	-3	1					28	$7/12$
7	Linear	-3	-2	-1	0	1	2	3				28	1
	Quadratic	5	0	-3	-4	-3	0	5				84	1
	Cubic	-1	1	1	0	-1	-1	1				6	$1/6$
	Quartic	3	-7	1	6	1	-7	3				154	$7/12$
8	Linear	-7	-5	-3	-1	1	3	5	7			168	2
	Quadratic	7	1	-3	-5	-5	-3	1	7			168	1
	Cubic	-7	5	7	3	-3	-7	-5	7			264	$2/3$
	Quartic	7	-13	-3	9	9	-3	-13	7			616	$7/12$
	Quintic	-7	23	-17	-15	15	17	-23	7			2184	$7/10$
9	Linear	-4	-3	-2	-1	0	1	2	3	4		60	1
	Quadratic	28	7	-8	-17	-20	-17	-8	7	28		2772	3
	Cubic	-14	7	13	9	0	-9	-13	-7	14		990	$5/6$
	Quartic	14	-21	-11	9	18	9	-11	-21	14		2002	$7/12$
	Quintic	-4	11	-4	-9	0	9	4	-11	4		468	$3/20$
10	Linear	-9	-7	-5	-3	-1	1	3	5	7	9	330	2
	Quadratic	6	2	-1	-3	-4	-4	-3	-1	2	6	132	$1/2$
	Cubic	-42	14	35	31	12	-12	-31	-35	-14	42	8580	$5/3$
	Quartic	18	-22	-17	3	18	18	3	-17	-22	18	2860	$5/12$
	Quintic	-6	14	-1	-11	-6	6	11	1	-14	6	780	$1/10$

Table B.11 Curves of Constant Power for Tests on Main Effects

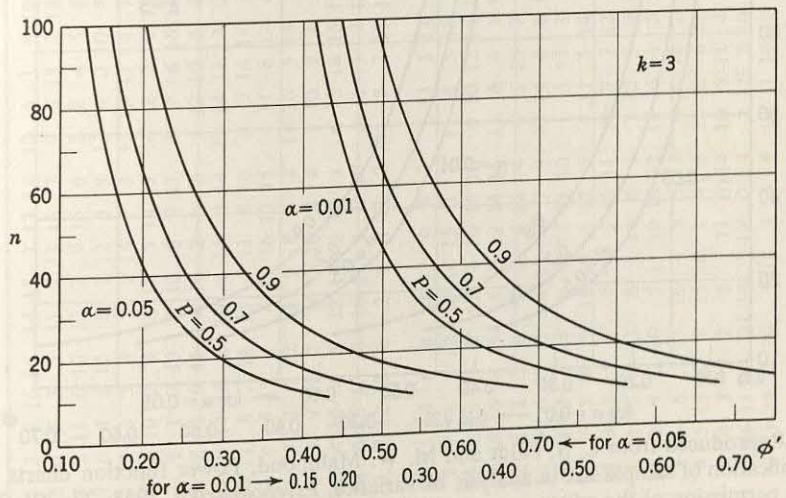
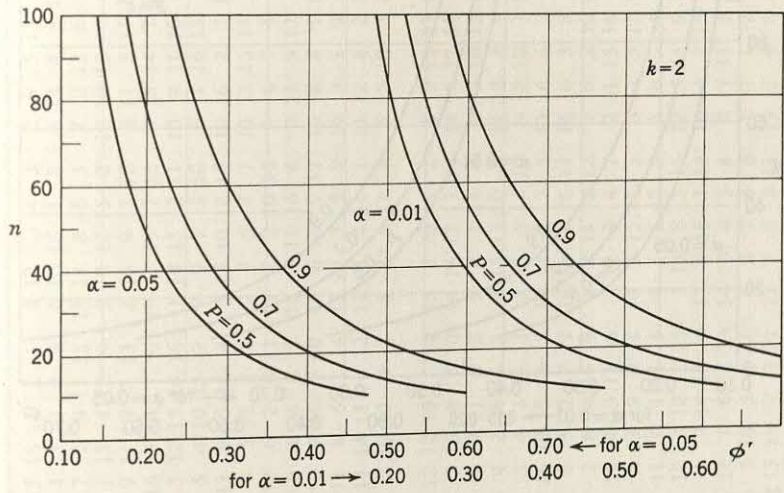
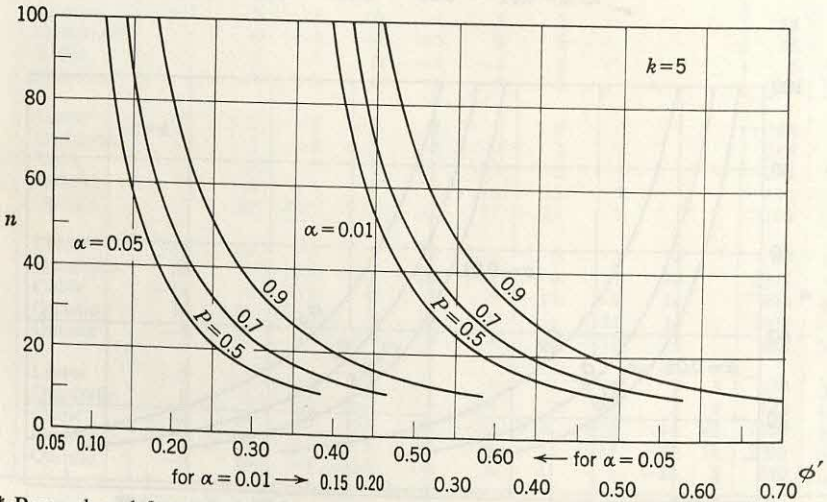
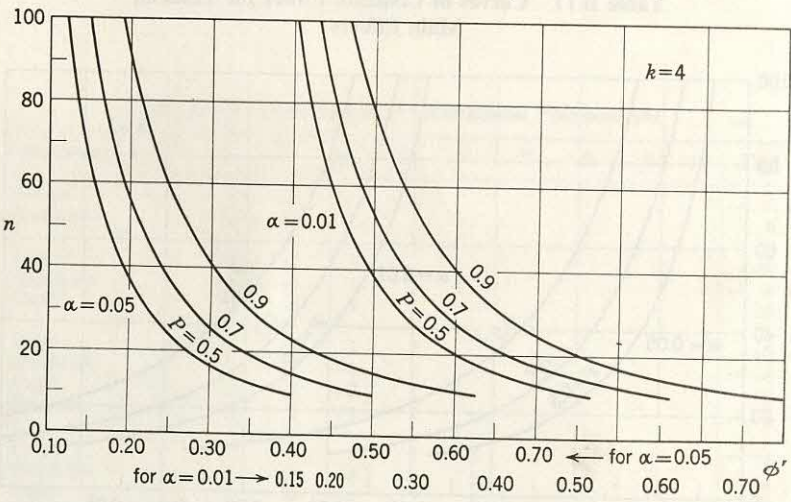


Table B.11 Curves of Constant Power for Tests on Main Effects (Continued)*



* Reproduced from L. S. Feldt and M. W. Mahmoud, Power function charts for specification of sample size in analysis of variance. *Psychometrika*, 1958, 23, 201-210, with permission of the editor.

Table B.12 Random Permutations of 16 Numbers

12	11	16	11	7	4	13	7	16	16	14	15	7	13	13	10	6	1	8	10	15	16	1	2	1	3	12	1	4	3	2	9	5	13	8	15	10	15	13	3
5	9	3	1	2	5	6	3	5	11	7	13	16	7	7	5	13	2	14	4	13	3	4	5	6	10	16	1	8	8	14	4	8	14	6	6	10	5	16	
13	15	2	5	6	8	2	11	6	10	13	2	9	15	6	2	8	4	5	8	7	2	9	15	15	14	11	8	6	5	7	2	12	6	2	13	2	13	15	10
16	4	5	16	9	9	16	4	8	15	8	5	12	3	9	13	4	3	10	4	1	15	4	12	9	9	8	12	15	2	12	10	13	10	1	12	14	8	7	5
2	2	13	1	11	11	12	13	5	3	5	8	14	7	3	15	13	11	4	7	9	11	7	1	13	4	16	9	13	12	4	11	3	11	10	8	8	9	12	4
10	12	10	7	14	15	14	10	12	4	3	12	5	14	12	16	2	2	9	15	16	3	14	7	6	10	1	13	7	16	9	15	14	7	11	14	9	4	8	15
8	14	15	14	12	7	1	1	7	13	7	10	6	9	11	9	7	6	16	11	3	12	13	16	16	1	14	15	2	11	6	4	11	16	6	4	11	7	14	6
3	16	3	6	16	6	8	3	9	11	4	16	2	6	5	1	12	10	6	9	5	4	10	13	7	7	9	7	3	10	6	1	16	15	9	16	5	1	4	11
1	1	8	9	2	16	10	14	14	2	10	1	15	5	14	12	14	12	3	2	13	9	5	14	4	5	15	10	12	15	10	16	2	3	13	10	7	2	1	12
15	3	12	8	3	10	11	8	2	14	15	4	11	8	16	8	9	14	14	1	8	5	15	3	8	2	7	5	8	7	13	7	15	14	4	1	4	14	10	7
6	6	1	10	13	13	9	5	15	1	6	13	16	12	15	4	11	8	12	6	12	6	2	8	11	13	2	2	10	14	1	6	7	4	7	9	1	12	16	1
4	10	4	4	15	3	15	2	1	12	1	9	8	1	8	14	16	5	13	5	6	1	8	6	2	16	5	11	16	6	3	13	9	12	5	5	3	16	6	8
7	8	7	15	10	5	3	9	11	7	9	3	10	4	4	3	10	9	1	3	10	7	16	5	10	15	6	14	9	13	16	8	6	1	15	11	15	11	14	
14	13	6	12	5	12	4	16	4	9	12	14	3	10	1	6	15	16	15	12	11	8	6	11	14	12	13	3	5	9	14	3	1	9	3	2	12	3	2	
11	5	14	2	4	14	6	15	10	8	16	11	1	2	10	5	1	15	7	13	14	14	11	10	12	11	4	4	14	1	11	12	8	2	16	7	13	5	9	13
9	7	11	13	8	1	7	12	13	6	2	6	4	11	2	11	3	7	11	16	2	10	12	9	3	8	3	6	11	4	5	5	10	5	12	3	16	6	2	9
1	10	1	15	16	10	11	4	6	4	14	12	9	8	7	5	13	3	13	3	12	12	10	2	14	15	3	10	9	2	8	9	12	13	10	3	16	8	9	8
3	3	5	3	6	7	7	8	5	12	10	14	15	1	2	11	4	5	15	9	7	15	8	15	10	4	16	12	13	14	14	8	11	14	2	10	6	11	6	6
10	8	11	11	7	16	12	15	9	9	12	6	16	15	16	9	10	12	16	15	9	9	11	6	2	9	8	1	15	4	6	3	16	12	14	1	12	16	11	2
9	2	4	7	1	3	15	10	14	11	3	3	12	14	15	13	6	4	1	16	4	2	16	4	8	8	15	13	10	8	11	13	4	1	6	7	15	12	14	11
6	16	9	4	8	12	8	2	8	6	2	9	8	16	4	6	5	15	7	8	3	4	3	14	12	2	5	6	14	10	9	12	10	7	13	8	2	1	8	9
2	12	10	9	12	14	9	5	13	5	4	13	7	9	12	14	8	8	11	16	16	10	15	8	3	10	14	2	7	3	15	6	6	3	9	2	3	9	1	13
12	1	15	5	15	6	4	12	4	14	9	5	14	11	10	1	3	13	3	5	8	5	2	7	7	16	11	14	16	13	16	10	3	8	7	9	1	13	7	12
13	6	8	13	11	3	1	10	7	11	15	6	12	5	7	11	1	14	4	16	15	14	13	11	6	3	12	3	4	7	2	11	15	16	8	4	10	2	10	4
14	5	3	16	4	13	10	13	15	3	7	8	3	13	11	8	7	7	12	7	15	14	13	11	6	3	12	3	4	7	2	11	15	16	8	4	10	2	10	4
8	15	2	14	9	1	1	7	3	13	8	1	13	6	3	3	15	9	9	12	6	13	9	9	1	11	6	15	11	5	3	14	5	11	15	6	9	6	3	15
11	13	16	12	10	15	2	9	2	16	5	11	10	10	12	15	16	14	5	2	11	8	1	13	5	1	1	5	12	9	16	12	15	7	4	4	5	13	15	16
16	9	6	10	14	5	6	6	7	15	16	4	5	5	13	4	9	16	2	6	10	3	12	12	15	7	13	8	3	16	12	15	14	9	12	11	8	7	15	5
7	14	12	8	3	4	5	16	12	8	15	10	1	4	6	16	2	6	11	1	1	7	14	10	4	13	4	16	1	15	10	4	8	15	5	13	4	10	13	7
4	4	13	2	5	8	14	3	1	1	6	7	11	9	14	10	8	11	4	13	14	1	6	5	13	14	10	4	5	1	4	7	13	5	11	12	5	4	2	1
5	7	14	1	2	2	16	11	16	2	13	2	4	2	1	2	12	2	10	14	5	16	5	1	16	6	9	7	6	6	1	2	1	2	3	14	7	14	16	14
15	11	7	13	11	9	13	14	11	10	1	16	2	3	8	14	1	10	6	10	13	11	7	3	11	12	7	9	8	11	13	1	9	6	1	16	11	3	4	10

Table B.12 Random Permutations of 16 Numbers (Continued)

12	13	16	6	12	10	5	12	1	10	4	6	6	1	4	5	7	13	2	10	16	11	15	9	2	12	4	10	10	1	9	6	15	15	1	13	1	1	16	2	
6	1	13	2	7	12	16	10	11	13	8	2	15	7	9	8	16	1	14	3	14	12	14	1	7	16	7	12	8	6	4	10	6	3	10	16	7	10	12	12	
10	14	11	12	4	15	13	16	9	3	7	16	11	8	3	3	12	2	3	4	8	10	13	6	5	5	9	5	1	7	3	12	11	2	15	3	16	16	6	7	
7	7	4	10	10	6	1	9	5	9	15	5	12	5	7	16	5	11	8	1	5	6	2	8	3	7	13	15	4	14	1	16	10	13	9	1	5	8	15	14	
2	16	10	8	11	14	4	3	8	15	14	9	2	16	1	12	6	14	4	13	13	8	8	12	10	4	2	6	14	11	6	8	4	6	4	10	6	7	1	16	
11	3	6	4	15	1	12	4	6	1	16	14	3	15	11	11	3	9	12	5	2	5	1	3	9	2	1	16	2	5	16	7	9	10	5	12	2	4	14	5	
9	11	1	3	6	2	15	14	7	4	9	10	16	2	10	7	10	16			6	4	10	5	14	8	8	8	6	13	5	1	13	7	12	9	9	13	9	8	
8	15	3	16	14	16	14	2	2	7	10	12	5	12	5	14	13	16	5	6	3	7	11	14	6	1	16	1	3	4	10	2	2	14	13	8	4	6	13	13	
16	5	14	14	8	3	7	15	10	12	3	13	14	4	2	1	14	8	6	12	7	3	16	16	16	6	6	11	9	12	14	5	7	12	2	6	10	2	11	1	
13	10	7	7	9	5	3	7	13	12	6	2	4	13	3	16	9	9	3	7	14	4	14	3	7	1	11	10	3	5	2	11	4	3	8	11	4	11	9	10	15
4	12	12	5	16	9	11	6	3	16	11	7	8	14	6	4	4	15	11		11	13	12	10	4	3	11	2	7	15	7	14	14	11	16	2	14	15	4	11	
5	9	8	9	3	13	8	8	14	11	13	1	9	10	12	10	11	10	9	8	15	15	4	11	12	15	5	14	12	10	8	11	1	9	3	14	8	3	7	3	
15	6	9	13	5	11	2	11	13	5	6	8	4	9	8	15	8	6	11	9	1	9	7	15	11	9	14	4	15	16	12	9	5	1	6	15	3	5	3	10	
3	4	15	15	13	7	6	5	4	14	1	11	7	11	13	6	1	15	13	15	9	1	5	13	15	10	12	13	11	8	13	13	12	5	14	11	15	14	8	9	
1	2	5	11	1	8	10	1	15	8	12	3	10	6	14	7	2	12	16	7	12	16	6	2	13	13	15	9	16	3	5	15	16	4	8	7	13	11	5	6	
14	8	2	1	2	4	9	7	16	2	5	15	1	13	15	13	15	5	1	2	10	2	9	4	8	14	3	7	13	9	2	3	8	16	7	5	12	12	2	4	
5	7	15	11	10	14	10	16	4	16	6	14	1	16	8	7	2	3	3	12	12	11	5	11	16	7	14	5	4	7	14	12	6	15	10	7	7	9	8	10	
8	6	1	2	5	1	11	13	13	15	16	4	15	4	3	12	12	1	4	7	10	5	11	12	7	4	11	3	5	8	2	10	2	2	15	9	3	11	5	12	
6	13	6	5	13	9	15	4	9	10	10	15	14	9	10	1	14	8	8	16	7	9	12	7	1	12	5	6	9	15	8	15	9	3	9	5	14	10	13	8	
16	4	3	16	2	8	9	8	7	7	5	13	16	1	7	5	11	2	9	3	3	6	1	8	14	9	7	8	8	9	1	1	5	13	12	8	2	7	10	6	
14	9	5	10	15	5	13	9	3	4	12	1	12	12	14	16	3	11	11	8	8	15	15	15	6	14	16	14	3	16	16	8	4	7	1	11	15	1	3		
10	10	10	3	4	12	1	3	2	11	7	11	6	3	11	11	10	4	5	1	6	1	4	1	13	3	1	2	14	12	3	3	1	7	1	10	10	8	7	14	
7	11	4	7	3	2	7	14	11	1	9	3	3	7	13	11	10	4	6	9	6	1	4	1	13	1	13	16	2	5	4	13	10	9	3	14	6	16	4	7	
1	2	15	7	9	4	14	15	8	8	4	9	13	13	6	8	15	9	1	14	9	8	8	9	5	10	6	15	11	13	11	7	16	6	3	16	2	11	1	7	
15	16	8	9	9	11	16	12	12	3	1	8	4	11	5	2	16	10	12	4	15	12	6	3	5	4	13	11	14	7	6	4	8	8	11	12	6	9	16		
2	3	13	1	8	10	5	10	5	12	2	16	5	8	1	9	5	12	16	6	16	14	9	5	2	8	3	7	7	1	11	14	3	12	2	16	13	5	16	2	
3	14	14	6	12	3	8	1	15	5	15	10	7	10	9	10	6	14	10	11	16	14	9	4	12	2	9	9	3	6	15	2	14	1	11	4	5	3	3	9	
9	15	11	14	6	7	6	2	6	6	14	2	8	2	4	13	8	5	15	5	5	16	14	4	12	2	9	9	3	6	15	2	14	1	11	4	5	3	3	9	
13	1	7	13	1	16	3	11	16	13	13	12	9	6	2	15	7	15	7	13	4	3	7	6	10	11	15	4	1	16	6	8	16	6	16	2	4	12	6	11	
12	8	16	8	11	15	2	6	14	2	11	5	10	15	12	3	1	13	13	10		2	4	16	16	15	13	2	10	13	10	9	7	11	5	5	6	9	14	12	15
4	5	9	4	14	13	12	5	1	14	8	6	11	14	15	6	9	16	2	2	1	10	3	2	11	15	10	1	16	2	10	4	15	11	13	9	15	13	15	13	
11	12	12	12	16	6	4	7	10	9	3	7	2	5	16	14	13	7	14	15	11	7	13	10	4	6	12	11	6	4	5	9	13	14	4	13	8	4	14	4	

Content References

- Alexander, H. W.: A general test for trend. *Psychol. Bull.*, 1946, 43, 533-557.
- Anderson, R. L.: Use of contingency tables in the analysis of consumer preference studies. *Biometrics*, 1959, 15, 582-590.
- and E. E. Houseman: Tables of orthogonal polynomial values extended to $N = 104$. *Res. Bull.* 297, April, 1942, Ames, Iowa.
- and T. A. Bancroft: *Statistical theory in research*. New York: McGraw-Hill, 1952.
- Anderson, T. W.: *Introduction to multivariate statistical analysis*. New York: Wiley, 1958.
- Aspin, A. A.: Tables for use in comparisons whose accuracy involves two variances separately estimated. *Biometrika*, 1949, 36, 290-293.
- Bargmann, R.: A study of dependence in multivariate normal analysis. *Mim. Ser.* 186, 1957, Inst. of Statist., Univer. of North Carolina.
- Bennett, C. A., and N. L. Franklin: *Statistical analysis in chemistry and the chemical industry*. New York: Wiley, 1954.
- Binder, A.: The choice of an error term in analysis of variance designs. *Psychometrika*, 1955, 20, 29-50.
- Bose, R. C.: Paired comparison designs for testing concordance between judges. *Biometrika*, 1956, 43, 113-121.
- , W. H. Clatworthy, and S. S. Shirkande: Tables of partially balanced designs with two associate classes. *Agric. exp. Station tech. Bull.* 107, 1954, Raleigh, N.C.
- and T. Shimamoto: Classification and analysis of partially balanced incomplete block designs with two associate classes. *J. Amer. statist. Ass.*, 1952, 47, 151-184.
- Box, G. E. P.: Problems in the analysis of growth and wear curves. *Biometrics*, 1950, 6, 362-389.
- : Non-normality and tests on variance. *Biometrika*, 1953, 40, 318-335.
- : Some theorems on quadratic forms applied in the study of analysis of variance problems. *Ann. math. Statist.*, 1954, 25, 290-302, 484-498.
- and J. S. Hunter: Multi-factor experimental designs for exploring response surfaces. *Ann. math. Statist.*, 1957, 28, 195-241.
- Bozovich, H., T. A. Bancroft, and H. O. Hartley: Power of analysis of variance procedures for certain incompletely specified models. *Ann. math. Statist.*, 1956, 27, 1017-1043.

- Cochran, W. G.: Some consequences when assumptions for the analysis of variance are not satisfied. *Biometrics*, 1947, 3, 22-38.
- : The comparison of percentages in matched samples. *Biometrika*, 1950, 37, 256-266.
- : Testing a linear relation among variances. *Biometrics*, 1951, 7, 17-32.
- : Analysis of covariance: its nature and use. *Biometrics*, 1957, 13, 261-281.
- and G. M. Cox: *Experimental designs*. (2d ed.) New York: Wiley, 1957.
- Cornfield, J., and J. W. Tukey: Average values of mean squares in factorials. *Ann. math. Statist.*, 1956, 27, 907-949.
- Cox, D. R.: *Planning of experiments*. New York: Wiley, 1958.
- Crump, S. L.: The present status of variance component analysis. *Biometrics*, 1951, 7, 1-16.
- Duncan, D. B.: Multiple range and multiple F tests. *Biometrics*, 1955, 11, 1-42.
- : Multiple range tests for correlated and heteroscedastic means. *Biometrics*, 1957, 13, 164-176.
- Dunnett, C. W.: A multiple comparison procedure for comparing several treatments with a control. *J. Amer. statist. Ass.*, 1955, 50, 1096-1121.
- Durbin, J.: Incomplete blocks in ranking experiments. *Brit. J. statist. Psychol.*, 1951, 4, 85-90.
- Eisenhart, C.: The assumptions underlying the analysis of variance. *Biometrics*, 1947, 3, 1-21.
- Federer, W. T.: *Experimental design*. New York: Macmillan, 1955.
- Feldt, L. S., and M. W. Mahmoud: Power function charts for specification of sample size in analysis of variance. *Psychometrika*, 1958, 23, 201-210.
- Fisher, R. A.: *The design of experiments*. (6th ed.) Edinburgh and London: Oliver & Boyd, 1951.
- and F. Yates: *Statistical tables for biological, agricultural, and medical research*. (4th ed.) Edinburgh and London: Oliver & Boyd, 1953.
- Gayen, A. K.: The distribution of Student's t in random samples of any size drawn from non-normal universes. *Biometrika*, 1949, 36, 353-369.
- Geisser, S.: A method for testing treatment effects in the presence of learning. *Biometrics*, 1959, 15, 389-395.
- and S. W. Greenhouse. An extension of Box's results on the use of the F distribution in multivariate analysis. *Ann. math. Statist.*, 1958, 29, 885-891.
- Gourlay, N.: F test bias for experimental designs of the Latin square type. *Psychometrika*, 1955, 20, 273-287.
- Grant, D. A.: The Latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, 45, 427-442.
- : Analysis of variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, 53, 141-154.
- Graybill, F. A., and R. B. Deal: Combining unbiased estimators. *Biometrics*, 1959, 15, 543-550.
- Green, B. F., and J. Tukey: Complex analysis of variance: general problems. *Psychometrika*, 1960, 25, 127-152.

- Greenhouse, S. W., and S. Geisser: On methods in the analysis of profile data. *Psychometrika*, 1959, 24, 95-112.
- Hughes, H. M., and M. B. Danford: *Repeated measurement designs, assuming equal variances and covariances*. Rep. 59-40, 1958, Air Univer., School of Aviation medicine, USAF, Randolph AFB, Texas.
- Jackson, J. E., and M. Fleckenstein: An evaluation of some statistical techniques used in the analysis of paired comparison data. *Biometrics*, 1957, 13, 51-64.
- Johnson, N. L., and B. L. Welch: Applications of the noncentral t distribution. *Biometrika*, 1940, 31, 362-389.
- Kemphorne, O.: *The design and analysis of experiments*. New York: Wiley, 1952.
- and W. T. Federer: The general theory of prime power lattice designs. *Biometrics*, 1948, 4, 54-79, 109-121.
- Kogan, L. S.: Variance designs in psychological research. *Psychol. Bull.*, 1953, 50, 1-40.
- McHugh, R. B., and D. S. Ellis: The post mortem testing of experimental comparisons. *Psychol. Bull.*, 1955, 52, 425-528.
- Morrison, M.: Fractional replication for mixed series. *Biometrics*, 1956, 12, 1-19.
- Odeh, R. E., and E. G. Olds: *Notes on the analysis of variance of logarithms of variances*. WADC tech. Note 59-82, 1959, Wright-Patterson AFB, Ohio.
- Olds, E. G., T. B. Mattson, and R. E. Odeh: *Notes on the use of transformations in the analysis of variance*. WADC tech. Rep. 56-308, 1956, Wright Air Development Center.
- Paull, A. E.: On preliminary tests for pooling mean squares in the analysis of variance. *Ann. math. Statist.*, 1950, 21, 539-556.
- Plackett, R. L., and J. P. Burman: The design of optimum multi-factorial experiments. *Biometrika*, 1946, 33, 305-325.
- Ramachandran, K. V.: A test on variances. *J. Am. statist. Ass.*, 1958, 53, 741-747.
- Rao, C. R.: *Advanced statistical methods in biometric research*. New York: Wiley, 1952.
- Rider, P. R., H. L. Harter, and M. D. Lum: *An elementary approach to the analysis of variance*. WADC tech. Rep. 56-20, 1956.
- Robson, D. S.: A simple method for construction of orthogonal polynomials when the independent variable is unequally spaced. *Biometrics*, 1959, 15, 187-191.
- Satterthwaite, F. E. An approximate distribution of estimates of variance components. *Biometrics Bull.*, 1946, 2, 110-114.
- Scheffé, H. A.: A method for judging all possible contrasts in the analysis of variance. *Biometrika*, 1953, 40, 87-104.
- : *The analysis of variance*. New York: Wiley, 1960.
- Siegel, S.: *Nonparametric statistics*. New York: McGraw-Hill, 1956.
- Steel, R. G. D.: A multiple comparison rank sum test: treatments versus control. *Biometrics*, 1959, 15, 560-572.
- Sutcliffe, J. P.: A general method of analysis of frequency data for multiple classification designs. *Psychol. Bull.*, 1957, 54, 134-137.

- Tukey, J. W.: One degree of freedom for nonadditivity. *Biometrics*, 1949, 5, 232-242.
- : Answer to query. *Biometrics*, 1955, 11, 111-113.
- : The comparative anatomy of transformations. *Ann. math. Statist.*, 1957, 28, 602-632.
- Walker, H., and J. Lev: *Statistical inference*. New York: Holt, 1953.
- Welch, B. L.: The generalization of Student's problem when several different population variances are involved. *Biometrika*, 1947, 34, 28-35.
- Wilk, M. B., and O. Kempthorne: Fixed, mixed, and random models. *J. Am. statist. Ass.*, 1955, 50, 1144-1167.
- and O. Kempthorne: Some aspects of the analysis of factorial experiments in a completely randomized design. *Ann. math. Statist.*, 1956, 27, 950-985.
- and ———: Nonadditivities in a Latin square design. *J. Am. statist. Ass.*, 1957, 52, 218-236.
- Youden, W. J., and J. S. Hunter: Partially replicated Latin squares. *Biometrics*, 1955, 11, 399-405.
- Zoellner, J. A., and O. Kempthorne: Incomplete block designs with blocks of size two plots. *Res. Bull.* 418, 1954, Agric. exp. Station, Iowa State College.

References to Experiments

- Aborn, M., H. Rubenstein, and T. D. Sterling: Sources of contextual constraints upon words in sentences. *J. exp. Psychol.*, 1959, 57, 171-180.
- Bamford, H. E., and M. L. Ritchie: Complex feedback displays in a man-machine system. *J. appl. Psychol.*, 1958, 42, 141-146.
- Birch, D., E. Burnstein, and R. A. Clark: Response strength as a function of hours of deprivation under a controlled maintenance schedule. *J. comp. physiol. Psychol.*, 1958, 51, 350-354.
- Briggs, G. E.: Retroactive inhibition as a function of the degree of original and interpolated learning. *J. exp. Psychol.*, 1957, 54, 60-67.
- , P. M. Fitts, and H. P. Bahrick: Learning and performance in a complex tracking task as a function of visual noise. *J. exp. Psychol.*, 1957, 53, 379-387.
- , ———, and ———: Effects of force and amplitude in a complex tracking task. *J. exp. Psychol.*, 1957, 54, 262-268.
- , ———, and ———: Transfer effects from a double integral tracking system. *J. exp. Psychol.*, 1958, 55, 135-142.
- Castenada, A., and L. P. Lipsett. Relation of stress and differential position habits to performance in motor learning. *J. exp. Psychol.*, 1959, 57, 25-30.
- Conrad, R.: Accuracy of recall using keyset and telephone dial, and the effect of a prefix digit. *J. appl. Psychol.*, 1958, 42, 285-288.
- Cotton, J. W., D. J. Lewis, and R. Metzger: Running behavior as a function of apparatus and of restriction of goal box activity. *J. comp. physiol. Psychol.*, 1958, 51, 336-341.
- Denenberg, V. H., and R. D. Myers: Learning and hormone activity. I: The effects of thyroid levels upon the acquisition and extinction of an operant response. *J. comp. physiol. Psychol.*, 1958, 51, 213-219.
- French, G. M.: A deficit associated with hypermobility in monkeys with lesions of the dorsolateral frontal granular cortex. *J. comp. physiol. Psychol.*, 1959, 52, 25-28.
- Gerathewohl, S. J., H. Strughold, and W. F. Taylor: The oculomotoric pattern of circular eye movements during increasing speed of rotation. *J. exp. Psychol.*, 1957, 53, 249-256.
- Gordon, N. B.: Learning a motor task under varied display conditions. *J. exp. Psychol.*, 1959, 57, 65-73.

- Jerison, H. J.: Effects of noise on human performance. *J. appl. Psychol.*, 1959, 43, 96-101.
- Leary, R. W.: Homogeneous and heterogeneous reward of monkeys. *J. comp. physiol. Psychol.*, 1958, 51, 706-710.
- Levison, B., and H. P. Zeigler: The effects of neonatal X irradiation upon learning in the rat. *J. comp. physiol. Psychol.*, 1959, 52, 53-55.
- Lewis, D. J., and J. W. Cotton: Partial reinforcement and nonresponse acquisition. *J. comp. physiol. Psychol.*, 1958, 51, 251-257.
- Martindale, R. L., and W. F. Lowe: Use of television for remote control: a preliminary study. *J. appl. Psychol.*, 1959, 43, 122-124.
- Meyer, D. R., and M. E. Noble: Summation of manifest anxiety and muscular tension. *J. exp. Psychol.*, 1958, 55, 599-602.
- Michels, K. M., W. Bevan, and H. C. Strassel: Discrimination learning and interdimensional transfer under conditions of systematically controlled visual experience. *J. comp. physiol. Psychol.*, 1958, 51, 778-781.
- Noble, C. E., and D. A. McNeely: The role of meaningfulness in paired-associate verbal learning. *J. exp. Psychol.*, 1957, 53, 16-22.
- Payne, R. B.: An extension of Hullian theory to response decrements resulting from drugs. *J. exp. Psychol.*, 1958, 55, 342-346.
- Prokasy, W. F., D. A. Grant, and N. Myers: Eyelid conditioning as a function of unconditioned stimulus intensity and intertrial interval. *J. exp. Psychol.*, 1958, 55, 242-246.
- Schrier, A. M.: Comparison of two methods of investigating the effect of amount of reward on performance. *J. comp. physiol. Psychol.*, 1958, 51, 725-731.
- Shore, M. F.: Perceptual efficiency as related to induced muscular effort and manifest anxiety. *J. exp. Psychol.*, 1958, 55, 179-183.
- Staats, A. W., et al.: Language conditioning of meaning to meaning using a semantic generalization paradigm. *J. exp. Psychol.*, 1957, 57, 187-192.
- Taylor, J. A.: Meaning, frequency, and visual duration threshold. *J. exp. Psychol.*, 1958, 55, 329-334.
- Wulff, J. J., and L. M. Stolurow: The role of class-descriptive cues in paired-associates learning. *J. exp. Psychol.*, 1957, 53, 199-206.
- Zimbardo, P. G., and N. E. Miller: Facilitation of exploration by hunger in rats. *J. comp. physiol. Psychol.*, 1958, 51, 43-46.

Index

- A posteriori test, 85
 - comparison of tests, 88
- Additivity test, 216–218
- Adjusted sum of squares, 226, 293–297
- Adjustment for block effects, 482
- Alias, 427
 - fractional replication, 449–450
 - incomplete blocks, 433
- Allowance, 91
- Anderson, R. L., 292, 624, 662
- Arcsin transformation, 221, 650
- Aspin, A. A., 37, 662
- Associate classes, 501–502
 - partially balanced designs, 501–503
- Balanced incomplete-block designs, 477–483
 - adjusted means, 477
 - computation, 486–490
 - efficiency factor, 479
 - intra- and interblock information, 481
 - repeated measures, 485
 - table, 479
- Bancroft, T. W., 292, 661
- Bargmann, R., 298, 661
- Bartlett test, 95
 - homogeneity of interactions, 329
- Behrens-Fisher distribution, 37
- Between-cell variation, 338
- Between-class regression, 585
- Blocks, construction, 380–381, 383–384, 425
- Bose, R. C., 503, 511, 661
- Box, G. E. P., 34, 62, 92, 110, 123, 219, 240, 305, 370, 538, 661
- Bozovich, H., 207, 661
- Carry-over effects, 300–301
- Ceiling effects and interaction, 257
- Chi square, combining independent tests, 43
 - equality of covariances, 369–373
 - interaction, 631
 - partition, 629–632
 - ranks, 625–626
- Chi-square distribution, 655
- Circulant designs, 505–506
- Cochran, W. G., 37, 324, 462, 478, 511, 515, 581, 585
- Cochran *C* test, 94, 654
- Cochran *Q* test, 139
- Combining intra- and interblock information, 481, 484
- Combining tests, 43
- Comparisons, 65
 - with a control, 89–92
 - control group, 627–629
 - factorial experiments, 207–211
 - individual, 207–211
 - orthogonal, 66
 - trend, 273, 278, 360–361
- Components of sum of squares, 67
 - orthogonal, 70
- Computational formulas, general, 278–281
- Computational symbols, definition, 230–232
 - three-factor experiment, 327–328
- Concomitant variates, 578
- Concordance coefficient, 137
- Confidence interval, means, 31
 - simultaneous, 88, 91
- Confounding, 379–380
 - $3 \times 2 \times 2$ factorial, 433–439
 - 3×3 factorial, 412–414, 418–422
 - $3 \times 3 \times 3$ factorial, 422–423
 - $3 \times 3 \times 3 \times 2$ factorial, 442–447
 - $5 \times 5 \times 5$ factorial, 423

- Congruence, 381
- Conservative test, repeated measures, 322
 - three-factor experiment, 340
- Contingency table, analysis of variance, 629–632
- Contrast (*see* Comparisons)
- Control, comparisons with, 89–92
- Control group, factorial experiment, 263–267
- Cornfield, J., 157, 195, 301, 318, 662
- Correlated error, 299–300
- Correlation, between-class, 604
 - linear, 75
 - within-class, 604, 615
- Covariance analysis, assumptions, 586–587
 - computation, 588–590, 600–603
 - factorial experiments, 595–599
 - illustrative applications, 605–606
 - individual comparisons, 592
 - multiple covariates, 621
 - purpose, 578–579
 - regression equations, 588
 - repeated measures, 606–618
 - single-factor experiment, 581–588
 - split-plot, 607
 - two covariates, 618–621
 - unequal cell frequencies, 594, 605
- Covariate, definition, 579
- Cox, G. M., 37, 324, 462, 478, 511, 515, 662
- Crossed classification, 185
- Deal, R. B., 461, 662
- Degrees of freedom, 52
 - F' statistic, 202
 - F'' statistic, 201
 - partition, 109
- Dichotomous data, 138, 629–632
- Doolittle algorithm, 294–295
- Duncan, D. B., 85, 662
- Dunnett, C. W., 89, 662
- Dunnett distribution, 651–652
 - tests against control, 90–91
- Dwyer algorithm, 296–297, 634
 - unequal cell frequencies, 294–297
- Effective error, 461
- Effective replications, 397
- Efficiency factor, 479, 490
- Eisenhart, C., 155, 662
- Equality of covariances, 369–373
- Error estimation, 150, 635–638
- Estimation, least-squares, 635–638
- Estimation, terminology, 6
 - unequal cell frequencies, 224–227
 - unequal group size, 374–378
- Expected value of mean squares, basic
 - principles, 156–158
 - definition, 154–155
 - derivation, 63, 195–199
 - nested factors, 187
 - partially hierarchical designs, 189
 - relation to F ratio, 160
 - repeated measures, 318, 320, 335–336, 348–350
 - three-factor experiment, 172
 - Tukey-Cornfield algorithm, 195–199
 - two-factor experiment, 155
- Experimental error, 58
 - definition, 150–151
 - estimation, 150–151
- Experimental unit, split-plot design, 192
- F distribution, 642–647
- F ratio, construction, 160–162, 199–202
- F' statistic, 200
- F'' statistic, 201
- F_{\max} distribution, 653
- F_{\max} test, 92–96
- Factor, definition, 141
 - fixed, 143
 - random, 144
- Factorial experiment, computation, 258
 - expected value of mean squares, 353
 - higher-order, 162–170
 - illustrative applications, 287–291
 - notation, 141–146, 228–230, 241
 - preliminary tests on model, 255–256
 - purpose, 140–141
 - repeated measures, 337–349
 - single control group, 263–267
 - terminology, 141–146
 - 2^k series, 283–287
 - unequal cell frequencies, 291–297
- Federer, W. T., 504, 662
- Fisher, R. A., 515, 521, 662
- Fleckenstein, M., 513, 663
- Fractional replication, 447–455
 - aliases, 450–454
 - computation, 452–453
 - construction, 451–452
 - incomplete blocks, 453–455
 - 3^k series, 453–455
 - 2^k series, 448–450
- Friedman test, 137, 624

- Geisser, S., 305, 538, 662, 663
- Generalized interaction, 427–433
fractional replication, 447–455
incomplete-block design, 433
- Grant, D. A., 365–367, 662
- Graybill, F. A., 461, 662
- Greco-Latin square, computation, 538
definition, 515
illustrative example, 576
plans, 577
repeated measures, 546–549
- Green, B. F., 207, 662
- Greenhouse, S. W., 305, 538, 663
- Harter, H. L., 221, 663
- Hartley F_{\max} test, 93, 653
- Heterogeneity, covariance, 123
variance, 36, 240–241
- Hierarchal design, 184–191
model, 187
partial, 199
- Homogeneity, covariance, 121
interaction with subjects, 329
regression, 587
variance, 33, 92, 239–240
- Hotelling's T^2 , 124, 345, 632–635
- Hypothesis testing terminology, 9
- Incomplete-block design, balanced, 379–398,
482–501
block size 2, 505–506, 511–513
combining intra- and interblock information, 482
distinct replications, 483–484
efficiency, 479
Latin square, 532
partial balance, 501–503, 506–510
repeated measures, 398
repetitions, 484
table, 479
Youden squares, 492–501
table, 495
- Individual comparisons, 207–211
computation, 238–239
trend components, 273–278
unequal sample size, 378
- Interactions, components, 386–388
computation, 279–280
confounding, 386
definition, 148–149
Interactions, generalized, 427
geometric interpretation, 178
higher-order, 169
modular notation, 384–394
number, 169–170
partial information, 392
profile, 179–180
scale of measurement, 220–221, 244–248
sign pattern, 400
simple, 178–183
subdivision, 384–386
three-factor components, 391–392
trend components, 212–213, 275
- Interblock error, 461
- Intercorrelation, average, 126
- Intrablock error, 460
- Intrablock treatment information, 510
- Jackson, J. E., 513, 663
- Kemphorne, O., 225, 241, 505, 521, 663
- Kruskal-Wallis H test, 622–623
- Latin square, balanced set, 518–519
balanced incomplete-block, 532
computation, 526, 531, 535
confounding, 521
conjugate, 518
cyclic permutation, 518
definition, 514
fractional replication, 524–527
illustrative applications, 542, 548, 553, 562,
563, 573
orthogonal, 515
plans, 536, 577
randomization procedure, 521
repeated measures, 538–575, 577
computation, 540–542, 558–560, 565,
570
order effects, 523–524
standard form, 516
uses, 521–524
- Lattice design, 456–477
adjusted means, 465
balanced, 457–460
computation, 463–468
rectangular, 504
repetitions, 470
simple, 463–465
two-dimensional, 504

- Lattice-square design, 468–471
 - adjusted means, 477
 - computation, 471–477
 - repetitions, 472
- Least-squares estimation, 63, 635–638
 - unequal cell frequencies, 224–227, 291–297
 - unequal group size, 375
- Lev, J., 623, 664
- Likelihood, 8
- Linear \times linear component, 275
- Linear estimation, 635–638
- Linked paired-comparison design, 511–513
- Lum, M. D., 223, 663
- Main effects, 146–148
 - computation, 278
 - definition, 146
 - estimation, 151
 - variance, 147
- Mann-Whitney U test, 623
- Maximum-likelihood principle, 9
- Mean square, defined, 153–154
 - expected value, 154
- Means tests, 24
- Measurement error, 124
- Missing data estimation, 281–283
- Model, additive, 219
 - mixed, 155, 299–301
 - repeated measures, 118–124
 - with interactions, 119–124
 - without interactions, 118–119
 - test for additivity, 216–218, 267–273
 - variance component, 155
- Modular arithmetic, 381–383
 - balanced sets, 383–384
- Modular notation, interactions, 393–394
- Monotonic transformations, 220
- Multifactor repeated measures, 347–353
- Multiple covariates, 618–621
- Nested effect, 184
 - computation, 258–263
 - interactions, 188
 - notation, 187
- Nested factors, 184–191
 - expected value of mean squares, 187
- “Never-pool” rule, 207
- Newman-Keuls test, 80
 - repeated measures, 309
 - sequence of tests, 83
- Nonadditivity test, 216–218, 267–273
- Noncentral F distribution, 162
- Nonlinearity test, 74
- Nonparametric analogues, 622–632
- Normal distribution, 640
- Normal equations, 225–227, 635–637
- Notation systems, 230, 241
- Olds, E. G., 219, 663
- Order effects, 301
- Ordered differences, 82
 - Duncan test, 85
 - Newman-Keuls test, 82
 - Tukey (a), 87
 - Tukey (b), 87
 - unequal sample size, 101–102
- Orthogonal polynomials, 72, 354
 - coefficients, 656
- Paired-comparison design, 511–513
- Partial balance, definition, 502
- Partially balanced design, 501–503
 - associate classes, 501–502
- Partially hierarchical experiment, 188
 - expected value of mean squares, 189
- Poisson variate transformation, 220
- Polynomials, orthogonal, 76–77
- Pooled variance, 151
- Pooling error terms, 202–207
 - example, 206
 - interactions with subjects, 321
 - nested factors, 261–263
 - sequential procedures, 204–206
- Power, 11–13, 104
 - noncentral F distribution, 162
 - sample size, 104
- Power charts, 657–658
- Preliminary tests on model, 202–207
 - factorial experiments, 255–256
- Profile of interactions, 179–180
- Proportional cell frequencies, 375
- Proportions, transformations, 221
- Protection level, 85
- q_r statistic, 77
- Quasi F ratio, construction, 199–202
 - degrees of freedom, 201–202
- Random permutations, 659–660
- Range, studentized, 77–80
 - transformations, 221–222

- Rao, C. R., 226, 292, 663
- Reduced sum of squares, 584
(See also Covariance analysis)
- Reliability, concordance coefficient, 137
definition, 126
estimation, 124–132
judges, 129–131
variation between, 129–131
- Repeated measures, 105
conservative tests, 306, 340
compared with usual tests, 322
contingency table, 623–627
covariance analysis, 606–618
covariance matrix, 369–374, 632–633
expected value of mean squares, 303, 318, 350
Greco-Latin square, 546–549
incomplete-block design, 398
Latin square, 538–575, 577
computation, 558–560
means, 39
model, 303
with interactions, 119–124
without interactions, 118–119
multifactor, 349–353
ordered differences, 309
simple effects, 323–324, 340
single-factor experiment, 105–111
computation, 111–116
three-factor experiment, 319–349
computation, 341–344
simple effects, 323, 340
 3×3 factorial, 415, 421–422
trend analysis, 353–361
unequal group size, 374–378
variance-covariance matrix, 312–314, 369–374, 632–633
- Regression, within-class, 581
- Repetitions, 470, 474
- Replicated experiment, 213–216
compared with nonreplicated, 215
- Response surface, 211, 278
components of interactions, 211–213
- Rider, P. R., 221, 663
- Sample size, determination, 104
power, 104
- Sampling terminology, 4
- Satterthwaite, F. E., 37, 201, 663
- Scale of measurement, choice, 244–247
transformations, 218–222
- Scheffé, H., 85, 88, 123, 162, 216, 663
- Sequence effects, control, 352–353
- Sequential pooling, 204–206
- Shimamoto, T., 503, 661
- Significance level, 10–13
- Simple effects, computations, 232, 237
definition, 174
expected value of mean squares, 178, 237
tests, 178
- Simultaneous confidence interval, 88, 91
- Spearman-Brown prediction formula, 127
- Split-plot design, 191–195
expected value of mean squares, 192
- Split-split plot design, 193–195
- Square-root algorithm, 296–297
- Standard error of difference, 27
- Steel, R. G. D., 627, 628, 663
- Structural model, 56
- Studentized range, 77
table, 648–649
- Subject \times treatment interaction, 318
- Surface, interaction, 182–183
- Sweep-out method, 227, 294–297
- Symmetry test, 373
- t statistic, computation, 31
relation to F statistic, 208
- t test, difference between means, 28, 641
- T^2 test, 124, 345, 632–635
- Three-factor factorial experiment, computation, 250, 258
expected value of mean squares, 172
repeated measures, 319–337
computation, 324–329
- Transformations, 218–222
additivity, 220
monotonic, 220
related to error variance, 219
- Treatment effect defined, 57
- Treatment \times subject interaction, 322
- Trend components, computation, 273–278
cubic, 364
differences, 360–361
examples, 365–369
factorial experiment, 211–213
interactions, 212–213, 275
linear, 73
main effects, 211–212
nonlinear, 74
orthogonal, 73
quadratic, 74, 363
regression equation, 76
tests, 70, 132, 353–361

- Tukey, J., 87, 157, 195, 207, 216, 219, 267, 301, 318, 664
- Tukey test for nonadditivity, 216-218, 267-273
- 2^k factorial experiment, computation, 283-287, 399-401
- specialized notation, 402
- 2³ factorial experiment, repeated measures, 409-412
- Unequal cell frequencies, 241-244, 374, 594, 605
- least-squares solution, 224-227
- Unequal sample size, proportional frequencies, 374-378
- single-factor experiment, 96-99
- unweighted-means analysis, 103-104
- Uniqueness, Latin-square, 575
- Unweighted-means analysis, 103-104, 241-244, 375-377
- Variance component model, 62
- Variance-covariance matrix, 115, 123, 344-345
- numerical example, 312-314
- symmetry test, 369-374
- Variance homogeneity tests, 92-96
- Variation, partition, 51
- Walker, H., 623, 664
- Welch, B. L., 37, 664
- Wilk, M. B., 521, 664
- Yates, F., 515, 662
- Yates notation, 393
- Youden square, 492-495
- computation, 496-501
- table, 495
- Zoellner, J. A., 505, 664



